

OM of: US-09-685-696-176 to: GenEmbl.* out_format : pfs
Date: Jan 25, 2002 11:20 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+2.p2n.model -DEV=slp
-Q/cgn2_1/USPTO_spool/US09685696/runat_25012002_145550_10645/app_query.fasta_1.648
-DB=GenEmbl -OFTM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-GAPEXT=7.000 -YGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFTM=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09685696_SCGN1_1.6369 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-685-696-176
Query length: 579
Database: GenEmbl.*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 2065.020000

| score_list: | Strd Orig | zScore | Escore | Len | Documentation |
|------------------|-----------|---------|----------|--------|----------------------------------|
| Sequence | | | | | |
| gb_pr:HSU76705 | + 2956.00 | 3401.17 | 3.6e-181 | 4155 | ! U76705 Human putative RNA bind |
| gb_pr:HSU7188 | + 2956.00 | 3401.11 | 3.6e-181 | 4181 | ! U97188 Homo sapiens putative R |
| gb_pr:AF117108 | + 2949.00 | 3393.95 | 1.0e-180 | 4171 | ! AF117108 Homo sapiens IGF-II m |
| gb_pr:AB046173 | + 2858.00 | 3293.94 | 2.6e-175 | 1740 | ! AB046173 Mus musculus mimp3 m |
| gb_pr:HS497021 | + 2956.50 | 2954.68 | 2.7e-156 | 129432 | ! AL023775 Human DNA sequence |
| gb_ov:AF064633 | + 2483.00 | 2858.92 | 5.7e-151 | 2697 | ! AF064633 Xenopus laevis Vgl RN |
| gb_ov:AF042353 | + 2483.00 | 2857.51 | 6.9e-151 | 3147 | ! AF042353 Xenopus laevis KH dom |
| gb_ov:AF059523 | + 2461.50 | 2836.11 | 1.1e-149 | 2161 | ! AF059523 Xenopus laevis RNA bi |
| gb_ov:AF064634 | + 2461.50 | 2835.13 | 1.2e-149 | 2405 | ! AF064634 Xenopus laevis Vgl RN |
| gb_ov:AF161270 | + 2261.50 | 2605.82 | 7.2e-137 | 2011 | ! AF161270 Danio rerio Vgl RNA b |
| gb_ov:AF026527 | + 2219.50 | 2557.27 | 3.6e-134 | 2021 | ! AF026527 Gallus gallus zipcode |
| gb_ov:AF061569 | + 2192.00 | 2524.65 | 2.4e-132 | 2223 | ! AF061569 Mus musculus coding r |
| gb_pr:AF117106 | + 2191.00 | 2523.88 | 2.3e-132 | 2130 | ! AF117106 Homo sapiens IGF-II m |
| gb_pr:AF198254 | + 2181.00 | 2511.32 | 1.3e-131 | 2381 | ! AF198254 Homo sapiens mRNA-bi |
| gb_pr:AF117107 | + 1950.50 | 2246.70 | 7.2e-117 | 2010 | ! AF117107 Homo sapiens IGF-II m |
| gb_pr:AF057352 | + 1868.00 | 2145.95 | 3.0e-111 | 3667 | ! AF057352 Homo sapiens hepatoc |
| gb_pr:AC020629 | + 1661.50 | 1874.09 | 4.1e-96 | 142971 | ! AC020629 Homo sapiens 12q BA |
| gb_htg:AC015706 | + 1556.50 | 1750.61 | 3.1e-89 | 182695 | ! AC015706 Homo sapiens chromo |
| gb_htg:AL1596177 | + 1556.50 | 1750.51 | 3.1e-89 | 184653 | ! AL1596177 Homo sapiens chromo |
| gb_htg:AL139152 | + 1549.50 | 1748.62 | 4.0e-89 | 93682 | ! AL139152 Homo sapiens chromo |
| gb_htg:AC024199 | + 1286.00 | 1437.98 | 8.0e-72 | 188273 | ! AC024199 Homo sapiens chromo |
| gb_htg:AC092447 | + 1281.00 | 1432.55 | 1.6e-71 | 181369 | ! AC092447 Homo sapiens chromo |
| gb_rv:MUSYBIC | + 1040.00 | 1193.55 | 3.3e-58 | 2437 | ! L35549 Mus musculus Y-box bind |
| gb_htg:AC069053 | + 1013.50 | 1133.78 | 7.1e-55 | 59747 | ! AC069053 Homo sapiens chromo |
| gb_in:AF241237 | + 835.00 | 954.38 | 7.0e-45 | 3188 | ! AF241237 Drosophila melanogast |
| gb_htg:AC012929 | + 742.50 | 823.11 | 1.4e-37 | 46580 | ! AC012929 Drosophila melanog |
| gb_in:AE003484 | + 742.50 | 805.71 | 1.3e-36 | 314087 | ! AE003484 Drosophila melanog |
| gb_htg:AC090245 | + 676.50 | 745.02 | 3.2e-33 | 57261 | ! AC090245 Homo sapiens chromo |
| gb_htg:AC026579 | + 616.50 | 674.19 | 2.8e-29 | 67823 | ! AC026579 Homo sapiens chromo |
| gb_htg:AC026579 | + 523.00 | 566.22 | 2.9e-23 | 67823 | ! AC026579 Homo sapiens chromo |
| gb_pr:AC021876 | + 480.00 | 512.61 | 2.8e-20 | 104668 | ! AC021876 Homo sapiens clone |
| gb_htg:AC023375 | + 480.00 | 508.85 | 4.6e-20 | 158105 | ! AC023375 Homo sapiens chromo |
| gb_rv:AC084407 | + 376.50 | 385.54 | 3.4e-13 | 239837 | ! AC084407 Mus Musculus Strain |
| gb_pr:AC005082 | + 367.50 | 378.30 | 8.5e-13 | 169739 | ! AC005082 Homo sapiens clone |
| gb_htg:AC090865 | + 305.00 | 317.07 | 2.2e-09 | 51119 | ! AC090865 Homo sapiens chromo |
| gb_htg:AC079780 | + 305.00 | 306.35 | 8.7e-09 | 165600 | ! AC079780 Homo sapiens chromo |
| gb_htg:AC023556 | + 298.00 | 298.63 | 2.3e-08 | 159122 | ! AC023556 Homo sapiens chromo |
| gb_in:CEM88 | + 276.50 | 288.23 | 8.9e-08 | 32723 | ! Z34802 Caenorhabditis elegans |
| gb_htg:AC016961 | + 257.50 | 249.66 | 1.2e-05 | 202643 | ! AC016961 Homo sapiens chromo |
| gb_in:DMUB | + 248.50 | 274.60 | 5.1e-07 | 4203 | ! X99340 D.melanogaster genomic |
| gb_htg:AC016961 | + 242.50 | 232.34 | 0.0001 | 202643 | ! AC016961 Homo sapiens chromo |

gb_pat:130334 + 241.00 271.11 8.0e-07 2384 ! I30334 Sequence 1 from pate
gb_pat:196058 + 241.00 271.11 8.0e-07 2384 ! 196058 Sequence 1 from pate
gb_pr:HSU05040 + 240.00 270.19 9.0e-07 2325 ! U05040 Human FUSE binding p
gb_pr:AF083898 + 238.00 269.93 9.3e-07 1857 ! AF083898 Homo sapiens RNA-b
seq_name: gb_pr:HSU76705
seq_documentation_block:
LOCUS HSU76705 4155 bp mRNA PRI 26-JAN-1999
DEFINITION Human putative RNA binding protein Koc1 mRNA, complete cds.
ACCESSION U76705
VERSION U76705.1 GI:4098296
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 4155)
AUTHORS Mueller-Pillasch,F., Lacher,U. and Wallrapp,C.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1996) Innere Medizin I, University of Ulm,
Robert-Koch-Str 8, Ulm 89081, Germany
FEATURES
source
location/Qualifiers
1..4155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p11.5"
/cell_line="Patau898s"
/tissue_type="pancreas"
251..1990
/function="putative RNA binding protein"
/note="Kh-domain containing protein; overexpressed in
cancer"
/codon_start=1
/product="Koc1"
/protein_id="AAD09223.1"
/db_xref="GI:4098297"
/translation="MNKLYIGNLSNAAPSDLESFKDAKIPVSGFLVKTGYAFVDC
PDESWALKAIEALSGKIELHGKPIEHSVPRKORIKLIQIRNIPHLQWELDSILV
QYGVSECEQVNTSETAVVNTYSSKQDQARQALDKLNGFQLETLKVAYIPDEMAA
QCNPLQPRGRGQRGSGSGVSKQPCDLPLRLVLTQFVGAIIKGEKATI
RNITKQTKSDIVHRKNGAAEKSTILSTPEGSAACKSILEIMHKPAQDIKFTEE
IPLTKIANNFVIGRIGRNLKIEQDTDTKITISPLQELTYNPERTIVKNGVE
TKCAEEETMKIRSYDNASMLQALHGLNLGLFPPTSGMPPPTSGPPSA
MTAPYPOFEQSETVHOFIPALSVGLIGKQGHIKOLSRFAGASIKTAPAEADAK
VRMVIITGPQAKQAQRIYKIKENFVSPKRVKLEAHIRVPSFRAGRVIGKGGK
TVNELQNLNSAEVYVPRDTPDENQVVVKITGHFYACQVQAKRIQELITQVKHQHQQ
KALOSGPPQSRKK"
BASE COUNT 1277 a 830 c 851 g 1181 t 16 others
ORIGIN
alignment_scores:
Quality: 2956.00 Length: 579
Ratio: 5.105 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-685-696-176 x HSU76705
Align seg 1/1 to: HSU76705 from: 1 to: 4155
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
|||||
251 ATGAACAAATGTATATCGAAACCTCAGGAGAACGCCGCCCTCGGA 300
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
301 CCTAGAAAGTATCTTCAGAGACGCCAAGATCCCGGTGTCGGGACCTTCC 350
|||||
34 euVallysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||

351 TGGTGAAGACTGGCTACGGTTCTGGTGGACTGCCCGGACGAGACTGGGCC 400
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
401 CTCAGAGGCATCGAGGCGCTTTCAGGTAAATAGAACTGCACGGGAAACC 450
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
451 CATAGAAGTTGAGCACTCGGTCCCAAAAGGCAAGGATTCGGAACCTTC 500
84 InIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
501 AGATACGAAATATCCCGCTCATTTACAGTGGAGGTGCTGGATGTTTA 550
101 LeuValGlnTrpGlyValValIleSerCysGluGlnValAsnThrAspSe 117
551 CTAGTCCAGTAGTGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
601 GGAACACTGCAGTTGTAATGTAACTATTTCCAGTAAGACCAAGCTAGAC 650
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
651 AAGCACTAGACAACTGAATGGATTTCAAGTTCAGTAAGATTTTACCTTGAAA 700
151 ValAlaTrpIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnCl 167
701 GTAGCCTATATCCCTGATGAATGGCCGCCCCAGCAAAACCCTTGCAGCA 750
167 nProArgLysArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
751 GCCCGAGGTCGCGGGGGCTTGGGACAGAGGGGCTCTCAAGGCGAGGGGT 800
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
801 CTCAGGATCGGTATCCAAAGCAGAAACCATGTGATTTGCCCTCTCGCGCTG 850
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
851 CTGGTTCCACCCCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 900
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
901 CATTCGGAAACATCAACCAAGACAGCCAGTCTTAAATCGATGCCACCGTA 950
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
951 AAGAAATCGCGGGCTGCTCAGAAAGTGGATTACTATCTCTCTACTCCT 1000
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG 267
1001 GAAGGCACCTCTCGCGCTTGTAAAGTCTATTTCTGGAGATTATGCATAAGGA 1050
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
1051 AGCTCAAGATATAAAATTCAGAGAGAGATCCCTTGAAGATTTTAGCTC 1100
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
1101 ATAATACTTTGTTGGAGCTTATTTGGTAAAGGAAGGAATAATCTTAAA 1150
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
1151 AAAATTGAGCAAGACACAGACACTAAAATCAGGATATCTCCATTGCAGGA 1200
317 uLeuThrLeuTrpAsnProGluArgThrIleThrValLysGlyAsnValG 334
1201 ATTGACGCTGTATAATCCAGAACGCACTATTACAGTTAAAGGCAATGTTG 1250
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
1251 AGACATGTGCCAAGCTGAGGAGGAGATCATGAAGAAATCAGGGAGTCT 1300

... 3

351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367
1301 TATGAAAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATTCCTGG 1350
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
1351 ATTAATCTGAACGCTTGGGTCTGTCCACCCACTTCAGGGATGCCAC 1400
384 toProThrSerGlyProProSerAlaMetThrProProfyrProGlnPhe 400
1401 CTCCACCTCAGGGCCCTTTCAGCATGACTCTCCCTACCCCGCAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
1451 GAGCAATCAGAAACGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgP 434
1501 CGGTGCCATCATCGGCAAGCAGGCGCAGACATCAAGCAGCTTCTCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
1551 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGAAGCAGCAGATGCTAAA 1600
451 ValArgMetValIleIleThrGlyProGluAlaGlnPheLysAlaGl 467
1601 CTGAGGATGGTGATTATCCTGAGCACACAGAGGCTCAGTTCAAGGCTCA 1650
467 nGlyArgIleThrGlyLysIleLysGluGluAsnPheValSerProLysG 484
1651 GGGAGAATTTATGGAAAAATTAAGAAAGAAACTTTGTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
1701 AAGAGTGAAACTTGAAGCTCATATCAGATGCCATCTTTGCTGCTGCTG 1750
501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1751 AGAGTTATTGGAAAAAGGAGCAAAACGGTGAATGAATTCAGAAATTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
1801 AAGTGCAGAACTTGTGTCCTCGTGCAGCAGACACCTGATGAGAAATGACC 1850
534 InValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
1851 AAGTGGTTGTCAAAATAACTGCTCACTTCTATGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLys 567
1901 AGAAAAATTCAGGAAATTTCTCAGCTCAGGTAAAGCAGCACCACAAACAGAA 1950
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
1951 GGCTCTGCAAAAGTGACCACCTCAGTCAAGACGGAAG 1987

seq_name: gb_pr:HSU97188

seq_documentation_block:

LOCUS HSU97188 4181 bp mRNA PRI 11-SEP-1998
DEFINITION Homo sapiens putative RNA binding protein KOC (koc) mRNA, complete cds.

ACCESSION U97188

VERSION U97188.1 GI:2105468

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 4181)

REFERENCE

AUTHORS Mueller-Pillasch,F., Lacher,U., Wallrapp,C., Michä,A.,

Zimmerhackl,F., Hameister,H., Varga,G., Friess,H., Buchler,M.,

Beger,H.G., Vila,M.R., Adler,G. and Gress,T.M.
 Cloning of a gene highly overexpressed in cancer coding for a novel
 KH-domain containing protein
 Oncogene 14 (22), 2729-2733 (1997)
 97322117
 REFERENCE 2 (bases 1 to 4181)
 Mueller-Pillasch,F., Lacher,U., Wallrapp,C., Micha,A.,
 Zimmerhackl,F., Hameister,H., Varga,G., Friess,H., Buchler,M.,
 Beger,H.G., Vila,M.R., Adler,G. and Gress,T.M.
 Direct Submission
 TITLE Submitted (11-APR-1997) Medizinische Klinik, Internal Medicine I,
 JOURNAL Robert-Koch-Str.8, Dlm 89081, Germany
 FEATURES Location/Qualifiers
 source 1..4181
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="p11.5"
 /cell_type="pancreatic cancer"
 /cell_line="Patu8988t"
 1..4181
 /gene="koc"
 251..1990
 /gene="koc"
 /function="possible proliferation of cancer tissues"
 /note="KH-domain containing protein overexpressed in
 cancer"
 /codon_start=1
 /product="putative RNA binding protein KOC"
 /protein_id="AAC35208.1"
 /db_xref="GI:2105469"
 /translation="MNKLYIGNLSENAAPSDLESIFDKAPIVSGPFLVKTYGAFVDC
 PYGWSWALKEALSGKTELHGKPIEVESHVPKROKRIQIQRNIPPHLOWEVLDSLLV
 ODGVECEQVNTDSEAVNVTVYSSKDOARQDLKNGFOLENTLKVAVIPDEMAA
 QONLQPRGRGLGQRGSSRQSGPSVSKQKPCDPLRLVLPYQFVGAIIKGKATI
 RNLTKQSKIDVHRKNAGAERSITILSTPEGTSAAKCSILEIMHKEAODIKFTEE
 IPLKILAHNNVPEGLIGKEGNLKKIEQDITKITISPLQELTYNPERTITVKGNE
 TCAKAEIEIMKIRIESYENDIASMNLQALHLPGLNLAIGLFPPTSGMPPTSGPPSA
 MTPYPOFEQSETTHQFIIPALSVGAIIGKOGHIIKOLSRFAGASIKIARAEAPDPA
 RMVLITNGPPEAQKAGRIYKIKIENFVSPKEVKLEAHIRVPSFAAGRVICKGCK
 TVNELNLSLSREAVVPRDQTPDENQVVKITGHFYACQVAQRKIQEILTVQVKHQHQ
 KALQSGPPQSRKK"
 BASE COUNT 1303 a 830 c 851 g 1181 t 16 others
 ORIGIN

alignment_scores:
 Quality: 2956.00 Length: 579
 Ratio: 5.105 Gaps: 0
 Percent similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x HSU97188 ..

Align seg 1/1 to: HSU97188 from: 1 to: 4181

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
 |||||
 251 ATGAACAACCTGATATCGGAACCTTCAGCGAAGCCGCCCTCGGA 300
 17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
 |||||
 301 CCTAGAAAGTATCTTCAAGACGCCAAGATCCCGGTGTCGGGACCTTC 350
 34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
 |||||
 351 TGGTGAAGACTGCTACCGCTTCGTGGACTGCCCGGACGAGAGCTGGCC 400
 51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
 |||||
 401 CTCAGGCCATCGAGCGCTTTTCAGTAAATAAGAACTCGCGGAAACC 450
 67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84

|||||
 451 CATAGAAGTTGACACTCGCTCCCAAAAGCAAGGATTTCGGAACCTTC 500
 84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
 |||||
 501 AGATACGAAATATCCGCCCTCATTTACAGTGGGAGGTGCTGGATAGTTA 550
 101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
 |||||
 551 CTAGTCCAGTATGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600
 117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
 |||||
 601 GGAACCTGCAGTTCTAAATGTAACCTATTCCAGTAAAGCAAGCTAGAC 650
 134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
 |||||
 651 AAGCACTAGACAACTGAATGGATTTTCAGTTAGAGAAATTTACACCTTAAA 700
 151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
 |||||
 701 GTAGCCTATATCCTGATGAATGGCGCCCGCCAGCAAAACCCCTTGCAGCA 750
 167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlys 184
 |||||
 751 GCCCCGAGGTGCGCGGGGCTTGGCAGAGGGGCTCTCAAGGCAGGGGT 800
 184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
 |||||
 801 CTCAGGATCCGTATCCAAGCAAAACCATGTGATTTGCCCTCGCGCTG 850
 201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
 |||||
 851 CTGGTCCCAACCAATTTGTTGGAGCCATCATAGGAAAAGAGGTGCCAC 900
 217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
 |||||
 901 CATTCGAACATCACCACCAACAGACCCAGTCTAAATCGATGTCCACGCTA 950
 234 yGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
 |||||
 951 AAAAAATGCGGGGCTGCTGAGAAGTCGATCTACTCTCTCTACTCCT 1000
 251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGl 267
 |||||
 1001 GAAGGCACCTCTCGGGCTTGTAGTCTATTCTGGAGATTATGCATAAGGA 1050
 267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
 |||||
 1051 AGCTCAAGATATAAAATTACAGAAGAGATCCCTTGAAGATTTTAGCTC 1100
 284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
 |||||
 1101 ATAATAACTTTGTGGACGCTTATTGGTAAAGAAAGAAATCTTTAAA 1150
 301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGl 317
 |||||
 1151 AAAATTGAGCAACACACAGACACTAAATCAGCATATCTCCATTTCAGGA 1200
 317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
 |||||
 1201 ATTGACGCTGTATAATCCAGACGCACATATTACAGTTAAAGCAATGCTG 1250
 334 LuThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
 |||||
 1251 AGACATGTGCAAGCTGAGGAGGAGATCATGAAGAAAAATCAGGGAGTCT 1300
 351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367
 |||||
 1301 TATGAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATTCCTGG 1350
 367 yLeuAsnLeuAlaLeuGlyLeuPheProThrSerGlyMetProp 384
 |||||

1351 ATTAAATCTGAACGCTTGGGTCTGTTCCACCCCACTTCAGGGATGCCAC 1400
384 roProThrSerGlyProProSerAlaMetThrProProTyProGlnPhe 400
1401 CTCACCTCAGGGCCCTTCAGCCATGACTCTCTCCCTACCCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
1451 GAGCAATCAGAAACGAGACTGTTTCATCAGTTTATCCAGAGCTATCAGT 1500
417 IGlyAlaIleIleGlyLysGlnGlnHisIleLysGlnLeuSerArgp 434
1501 CGTCCCATCATCGCAGCAGGCGCACATCAAGCAGCTTCTCCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
1551 TTGCTGGAGCTTCAATTAGATTGCTCCAGCGGAGCACCAGATGCTTAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaGl 467
1601 GTGAGGATGGTGATTATCCTGACGACACAGAGGCTCAGTTCAGGCTCA 1650
467 nGlyArgIleThrGlyLysIleLysGluGluAsnPheValSerProLysG 484
1651 GGAAGAATTTATGGAAAATTTAAGAAGAAACTTTGTTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaIaGly 500
1701 AAGAGTCAAACTTGAACCTCATATCAGAGTCCCATCTTCTGCTGCTGC 1750
501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1751 AGAGTTATTGGAAAAGGAGGCAAAACCGTGAATGAACCTTCAGAAATTTGTC 1800
517 rSerAlaGluValValValProArgAspGlnThrProAspGluAsnAspG 534
1801 AAGTGCAGAGTTGTTGTCCTCGTACACAGACCTGATGAGAAATGACC 1850
534 lnValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
1851 AAGTGTGTCAAAATACTGGTCACCTCTATGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLy 567
1901 AGAAAAATTACAGAAATTTCTGACTCAGGTAAGCAGCACCACCAACAGAA 1950
567 salAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
1951 GGCTCTGCAAGTGGACCACCTCAGTCAAGACGGAAG 1987

seq_name: gb_pr:AF117108

seq_documentation_block:
LOCUS AF117108 4171 bp mRNA PRI 26-JAN-1999
DEFINITION Homo sapiens IGF-II mRNA-binding protein 3 (IMP-3) mRNA, complete cds.

ACCESSION AF117108
VERSION AF117108.1 GI:4191611

KEYWORDS

SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4171)
Nilsen, J., Christensen, J., Lykke-Andersen, J., Johnsen, A.H.,
Wewer, U.M. and Nilsen, F.C.

A family of insulin-like growth factor II mRNA-binding proteins
represses translation in late development
Mol. Cell. Biol. 19 (2), 1262-1270 (1999)

JOURNAL MEDLINE 99108099

REFERENCE 2 (bases 1 to 4171)
Nilsen, J., Christensen, J., Lykke-Andersen, J., Johnsen, A.H.,
Wewer, U.M. and Nilsen, F.C.

AUTHORS

TITLE Direct Submission
JOURNAL Submitted (30-DEC-1998) Institute of Molecular Biology, University of Copenhagen, Soelvgade 83H, Copenhagen DK-1307, Denmark
FEATURES
Location/Qualifiers
source 1..4171
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1..4171
/gene="IMP-3"
CDS 267..2006
/gene="IMP-3"
/note="binds multiple sites in IGF-II mRNA 5'UTR; identical to KH-domain containing protein overexpressed in cancer (koc)"
/codon_start=1
/product="IGF-II mRNA-binding protein 3"
/protein_id="AAD09828.1"
/db_xref="GI:4191612"
/translation="MNKLYIGNLSENAAPSDLESIFPKDAKIPVSGPPLVKGTGAFVDC PDESWALKATEALSGLKIEUHGKPIEVEHSVPKQRIKQIRINPHLOWEVLDSLIV QYGVESCEQVNTDSTAVVNTYSSKQARQALDKLNGFQLENFTLKVAY IPDEMAA QONPQQPRGRLGORGSSROGSPGVSQKQPCDLPRLVPTQFVGATIGKEGATI RNITKQTSKIDVHRKENAGAEKSIITLSTPEGTSNACKSILEIMHKEAQDIKFTTEE IPUKILAHNNFVRLIGKEGRNLUKKIEQDITDITITISPLQELTLYINPRTITVKNVE TCABEEETMKTRESYENDIASMNLQAHILPGLNINLALGFLPPTSGMPPTSGPPSA MTPPYPPQFQSETETVHLFTIPALSVAIIIGKQGHKQSRFAGASIKIAPAPAPGAK VRMVIITGPPEAQFKAQRIYKIKIENEFVSPKEVKELEAHIRVPSFAAGRVIGKGGK TVNELONLSAEVVVPDQTPDENDQVVKITGHFYACQVQAKRIQIELLTQVKHQHQQ KALQSGPPQSRRK"
BASE COUNT 1292 a 822 c 853 g 1199 t 5 others
ORIGIN
alignment_scores:
Quality: 2949.00 Length: 579
Ratio: 5.102 Gaps: 0
Percent Similarity: 99.827 Percent Identity: 99.827
alignment_block:
US-09-685-696-176 x AF117108 ..
Align seg 1/1 to: AF117108 from: 1 to: 4171
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
267 ATGAACAACTGTATATCGAAACCTCAGCGAGAACGCCGCCCTCGGA 316
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
317 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGTCGGGACCCCTCC 366
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
367 TGGTGAAGACTGGCTACCGCTGCTGGACTCCCGGACGAGCTGGGCC 416
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
417 CTCAAGGCCATCGAGGCGCTTTCAGGTAATAATAGAACTGCACGGGAACC 466
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
467 CATAGAAGTTGAGCACTCGGTCCCAAAAAGGCAAGGATTCGGAACATTC 516
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
517 AGATACGAAATATCCCGCTCATTTACAGTGGAGGTGCTGGATAGTTTA 566
101 LeuValGlnTyrGlyValValIcuSerCysGluGlnValAsnThrAspSe 117
567 CTAGTCCAGTATGAGTGTGAGAGCTGTGAGCAAGTGAACACCTGACTC 616
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||


```
617 GGAACCTGCAGTTGTAATGTAACTATTCCAGTAAGGACCAAGCTAGAC 666
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
667 AAGCAGTACAGAACTCAATGGATTTCAGTTAGAGAATTTCCACTTGAAA 716
151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnGln 167
|||||
717 GTAGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGACGAC 766
167 nProArgGlyValArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
|||||
767 GCCCGAGGTGCGCGGGGCTTGGCAGAGGGCTCCTCAAGGCGGGGT 816
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||
817 CTCCAGGATCCGTATCCAGCAGAAACCATGTGATTTGCCCTCGCGCCTG 866
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaIth 217
|||||
867 CTGGTTCCCAACCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 916
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
|||||
917 CATTCGGAACATACCAACAGACCCAGTCTAAATCGATGTCACCGTA 966
234 ysGluAsnAlaGlyAlaAlaGlyLysSerIleThrIleLeuSerThrPro 250
|||||
967 AAGAAATGCGGGGCTGCTGAGAAGTCGATTACTATCCTCTACTCCT 1016
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGln 267
|||||
1017 GAAGGCACCTCTCGCGCTTGTAAGTCTATTCTGGAGATTATGCATAAGGA 1066
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
|||||
1067 AGCTCAAGATATAAAATTCACAGAGAGATCCCTTGAAGATTTTAGCTC 1116
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
1117 ATAATACTTTGTTGGAGCTCTTATTTGGTAAAGAGGAAGAAATCTTAAA 1166
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGln 317
|||||
1167 AAAATTGAGCAAGACACAGACACTAAAATCACGATATCTCCATTGCAGGA 1216
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
1217 ATTGACGCTGTATTAATCCAGAACGCACATATTACAGTTTAAAGGCAATGTTG 1266
334 LuThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
1267 AGACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATCAGGAGTCT 1316
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGln 367
1317 TATGAAATGATATTGCTTCTATCAATCTTCAAGCACATTTAATTCCTGG 1366
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
1367 ATTAATCTGAACCCCTGGGTCTGTTCCCAACCCACTTCAGGGATGCGAC 1416
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
1417 CTCCCACCTCAGGCGCCCTTTCAGCCATGACTCTCTCCCTACCCGAGTTT 1466
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1467 GAGCAATCAGAAACGAGAGACTGTTTCATCTGTTTATCCAGCTCTATCAGT 1516
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgp 434
|||||
1517 CGGTGCCATCATCGCAAGCAGGCGCCAGCATCAAGCAGCTTCTCGCT 1566
```

```
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1567 TTGCTGCAGCTTCAATTAAAGATTGCTCCAGCAGAAGCACCAGATGCTAAA 1616
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaGln 467
|||||
1617 GTGAGAGTGTGATTATCACTGGACACCACAGAGGCTCAGTTCAAGGCTCA 1666
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1667 GGGAGGAATTTATGAAAAATTAAGAAGAAACTTTGTTAGTCCTRAAG 1716
484 LuGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1717 AAGAGGTGAAACTTGAAGTCTATATCAGAGTGCATCTCTTGTGCTGTGGC 1766
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1767 AGAGTTATTGAAAAGAGGCAAAACGGTGAATGAACCTCAGAATTTGTC 1816
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1817 AAGTGCAGAGTTGTTGTCCTCTGTGACACACACCTGATGAGATGACC 1866
534 InValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1867 AAGTGGTTGTCAAAATAACTGGTGCATCTTATGCTTGCACAGTTGCCAG 1916
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGln 567
|||||
1917 AGAAAAATTCAGGAATTTCTGACTCAGGTAAGCAGCACCACCAACAGAA 1966
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1967 GGCTCTCAAGTGGACCACTCAGTCAAGACGGAAG 2003
seq_name: gb_ro:AB046173
seq_documentation_block:
LOCUS AB046173 1740 bp mRNA ROD 23-DEC-2000
DEFINITION Mus musculus mImp3 mRNA for igf2 mRNA-binding protein 3, complete cds.
ACCESSION AB046173
VERSION AB046173.1 GI:119333383
KEYWORDS
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Mori,H., Sakakibara,S., Imai,T., Nakamura,Y., Iijima,T., Suzuki,A.,
Yuasa,Y., Takeda,M. and Okano,H.
TITLE Expression of mouse igf2 mRNA-binding protein 3 and its
implications for the developing central nervous system
JOURNAL J. Neurosci. Res. (2001) In press
REFERENCE 2 (bases 1 to 1740)
AUTHORS Mori,H., Okano,H., Sakakibara,S. and Imai,T.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2000) Hiroshi Mori, Osaka Univ. Graduate School
of Medicine, Division of Neuroanatomy (D12); 2-2 Yamadaoka, Suita,
Osaka 565-0871, Japan (E-mail:hmori@nana.med.osaka-u.ac.jp,
Tel:81-6-6879-3581, Fax:81-6-6879-3589)
FEATURES
Source
1..1740
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1..1740
/genes="mimp3"
CDS
1..1740
/genes="mimp3"
/codon_start=1
/product="igf2 mRNA-binding protein 3"
```

```
/protein_id="BAB19755.1"
/db_xref="GI:11933384"
/translation="MNKLYIGNLSDHAGPADLESFVKDAKIPVAGPFLVKVTGYAFVDC
PDEGWALKAEIALSGKMLHGKPMVEHVSVPKRRIRKLIQINIPPHLOWEVLDSLLV
QYGVESCEQVNTDSTAVVNTYSSKDAQALDKLNGFQLENFTLKVAYIPDETA
QNPSPQLGRGPGGSGRSRASPQSGVSKQPCDPLRLVPTQVGAIIIGEGATI
RNITKQTSKIDVHRKENTGAAPKSITILSTPEGTSAAKSIIEIMHKEAODIKFTEE
IPKILAHNFVRLIGKEGRNLKITEQDTKITISPLQELTLYNPRTITVKGSGVE
TCAKABEEMKKIKRESYENDIASMNLQAHILPGLNLAHLGFPPTSGMPPTSGPST
LTPPYQFQSESETVHLFIPLALSVCAIIGKOGHOIKOLSRFAGASIKITAPAEAPDAK
VMVLIITGPEPAQKAGRIYKIKIENEFVSPKEEVKLEAHIRVPSFAAGRVIGKGGK
TNELSGSPSSRRK"
BASE COUNT      500 a      431 c      442 g      367 t
ORIGIN
```

```
alignment_scores:
  Quality: 2858.00      Length: 579
  Ratio: 4.988          Gaps: 0
  Percent Similarity: 98.964      Percent Identity: 96.373

alignment_block:
US-09-685-696-176 x AB046173 ..
Align seg 1/1 to: AB046173 from: 1 to: 1740
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaLaProSerAs 17
|||||
1 ATGAACAAATTTGATCATCGGAACTTACGACACCGCCGCCCGCGGA 50
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
51 CCTGAAAGTCTTCAAGGACGCTAAGATCCCGTGGCGGCCCTTCC 100
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
101 TGGTGAAGACGGGTACGCTTCGTGGACTGCCCGGACGAGGCGTGGGCC 150
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
151 CTCAGGCCATCGAGGCGCTTCAGGTAAATGGAACATACATGGAACACC 200
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
201 GATGGAAGTTGAGCACTCGGTCTTAAACGCGCAGAGGATTCGTAACATC 250
84 InileArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
251 AGATACGAAATATCCGCCCCACCTTACAATGGGAGGTGCTGGATAGTTTA 300
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
301 CTAGTCAGATGAGTGGTGGAGAGCTGTGAGCAAGTGAACACGGATTC 350
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
351 GGAACGCGCAGTTGTAATGTAACCTATTCCAGTAAGGACCAAGCTAGAC 400
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
401 AAGCCTTAGACAAACTGAATGGATTCCAGTTAGAGAACTTCACCTTGAAG 450
151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnGln 167
|||||
451 GTTCCTCATATCCAGATGAACATGCTGCCACGCAAAATCCCTCGCCACA 500
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlys 184
|||||
501 GCTCGGGGGCGCGGGCGGCGAGGCGGCGCTCATCCAGGCGAGCGGT 550
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||
```

```
551 CTCAGGATCGGTGTCCAAAGCAGAAACCCCTGTGACCTGCCACTGCCCTG 600
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
|||||
601 CTGGTTCACCACCGATTGTTGGAGCCATTATAGGAAAGAGGTGCTTAC 650
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
|||||
651 CATTCGCAACATCAAAAGCAGACCCAGCTCTAAAATCGATGCCATCCTA 700
234 ysGluAsnAlaGlyAlaGluLysSerIleThrIleLeuSerThrPro 250
|||||
701 AGGAGATACAGGGCGCGGAGAGTCCATTACTATCTCTTACCCT 750
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG1 267
|||||
751 GAAGCACCTCTGCAGCTTGTAACTTATCTGGAGATTATGCATAAGGA 800
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
|||||
801 AGCTCAAGATATAAAATTCACAGAGGAGATTCCCTTGAAGATCTTAGCTC 850
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
851 ACAATAACTTTGTAGGCGCTCTCATTTGTTAAAGAGGAAGAAACCTTAA 900
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG1 317
|||||
901 AAAATCGAGCAAGCACGACACATAAATCACATATCTCCATTCACGAGA 950
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
951 ACTGACGCTGTACAATCGGAACGACCATTTACAGTGAAGGACGAGTGTG 1000
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1001 AGAGCTGTGCCAAGCGGAGGAGGAAATAAATGAAGAGATCAGGGAGTCT 1050
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG1 367
|||||
1051 TATGAAATGATATTGCTTCCATGAACTTCAAGCACATTTAATCCCTGG 1100
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
|||||
1101 ATTTAAATCTGAATGCTTGGTGTGTTCCACCCACGTCAGGATGCCAC 1150
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1151 CTCCACCTCAGGGCCCCCTTCAACCCCTGACTCCTCCCTACCCACAATTT 1200
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1201 GAGCAATCAGAGCGGAGACTGTGCATCTGTTTATCCCGCCCTGTCCGT 1250
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgP 434
|||||
1251 TGGCGCCATCATTTGGCAAGCAGGCGCAACACATCAACAGAGCTTCTCGCT 1300
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1301 TTGGGGAGGCTTCGATTAAAGATCGTCCAGCCGAGGACCATGATGCTAAA 1350
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG1 467
|||||
1351 GTGGGATGTTGATTATCACTGGACACCCAGAGGCTCAGTTCAAGGCTCA 1400
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1401 GGGGAAGAAATTTATGGAAAAATTAAGAAAGAAACCTTTTGTAGTCTTAAAG 1450
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1451 AAGAGGTGAAACTTTGAAGCTCACATCAGAGTCCGCTCTTTTGTGCTGGC 1500
```

501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1501 AGAGTTATTGGGAAGGAGGCAAAACGGTGAATGAGCTCCAGAGCTTTATC 1550
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
1551 AAGTGTCAAGTTGTCTGCCCGGTGACCACACACCTGTATCAGATGATC 1600
534 InValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
1601 AAGTAGTTGTCAAAATAACTGCGCCACTCTATGCTTGCCAGGTGCCAG 1650
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLy 567
1651 AGGAANATTCAAGAAATTCGTGACTCAGGTAAAGCAGCACCCAGCAGAA 1700
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
1701 AGCTCTGCAGAGTGGACCACTCAGTCAAGCGGAAG 1737

```

/db_xref="taxon:9606"
/chromosome="6"
/map="q26-27"
/clone="RP3-497J21"
/clone_lib="RPCI-3"
complement(1..62)
/note="AluX repeat: matches 62..1 of consensus"
27..424
/note="match: STS AL022447"
63..231
/note="MER20 repeat: matches 53..218 of consensus"
433..591
/note="3 copies 53 mer 86% conserved"
complement(1513..1915)
/note="MLTIB repeat: matches 382..2 of consensus"
complement(2671..2713)
/note="MLTIB repeat: matches 47..3 of consensus"
complement(3291..3586)
/note="AluSg repeat: matches 298..1 of consensus"
4989..5126
/note="L1MB6 repeat: matches 1..141 of consensus"
5137..5648
/note="L1MB5 repeat: matches 382..923 of consensus"
complement(7362..7463)
/note="MIR repeat: matches 194..87 of consensus"
8372..8668
/note="AluX repeat: matches 2..300 of consensus"
8670..8703
/note="17 copies 2 mer aa 82% conserved"
14252..14295
/note="22 copies 2 mer gt 98% conserved"
14254..14301
/note="3 copies 16 mer 94% conserved"
14254..14297
/note="11 copies 4 mer ggtg 98% conserved"
14441..14481
/note="MIR2 repeat: matches 99..139 of consensus"
complement(16259..16324)
/note="MIR repeat: matches 157..92 of consensus"
complement(18532..18711)
/note="MIR repeat: matches 218..29 of consensus"
20727..21002
/note="AluX repeat: matches 2..296 of consensus"
22965..23004
/note="10 copies 4 mer caca 83% conserved"
complement(24989..25171)
/note="L1MB3 repeat: matches 931..738 of consensus"
complement(25204..25504)
/note="AluSp repeat: matches 303..1 of consensus"
complement(25761..26017)
/note="AluY repeat: matches 301..45 of consensus"
complement(27102..27194)
/note="MIR repeat: matches 142..50 of consensus"
30090..30204
/note="5 copies 23 mer 94% conserved"
complement(31383..31419)
/note="MIR2 repeat: matches 113..76 of consensus"
31912..32368
/note="match: STS S53188; part 1 of genomic marker D6S193"
32276..33267
/note="31 copies 32 mer 91% conserved"
33141
/note="tandem repeat"
33204
/note="tandem repeat"
33222..33396
/note="tandem repeat and single clone region"
33268..33555
/note="9 copies 32 mer 88% conserved"
33288..33539
/note="17 copies 16 mer 89% conserved"
33473..33793
/note="match: STS S53190; part 2 of genomic marker D6S193"

```



```
79931 TGTCACCCGTAAGAAATGACGGGCTGCTGAGAAAGTCGATTACTATCC 79980
|||||
247 euSerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 263
|||||
79981 TCCTACTCTGAAGCACCCTGCGGCTGTGAAGTCTATCTGGAGATT 80030
|||||
264 MethLysGluAlaGlnAspIleLysPheThrGluGluIleProLeu 280
|||||
80031 ATCGTAAGGAAACTCAAGATGTAATAATCACAGAAGAGATCCCTTGAA 80080
|||||
280 sIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGly 297
|||||
80081 GATTTAGCTCAAAATAACTTTCTGGCGCTCTATTGGTAAAGAAGAA 80130
|||||
297 rGAsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSer 313
|||||
80131 GAATCTTAAANAATTTGACAGACACAGACACTAAATATCAGATATCT 80180
|||||
314 ProLeuGlnGluLeuThrLeuTyAsnProGluArgThrIleThrVally 330
|||||
80181 CCATTGCAGGAACGTGTCAGTATATATCCAGAACACACACATTACAGTAA 80230
|||||
330 sGlyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysI 347
|||||
80231 AGGCAATGTTGAGACATGTGCCAAGCTGAGGAAGAGATCATGAAGAAA 80280
|||||
347 leArgGluSerTyGluAsnAspIleAlaSerMetAsnLeu.GlnAlaHi 363
|||||
80281 TCAGGAGCGCTTAGGAAATGATACATCTCTATG...ATTCAAGCACA 80327
|||||
363 sLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 380
|||||
80328 TTTAATTCCTGGATTAATCTGACGCTTGGCTCTCTCCACCCACTT 80377
|||||
380 erGlyMetProProProThrSerGlyProProSerAlaMetThrProPro 396
|||||
80378 CAGGGATGCCACCTCCACCTCAGGCGCCCTTCAGCCATGACTCTCC 80427
|||||
397 TyProGlnPheGluGlnSerGluThrGluThrValHisGlnPheIlePr 413
|||||
80428 TCCCGCAGTCTGAGCAATCAGAAATCGAGACTGTCATCTGTTATCCC 80477
|||||
413 oAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysG 430
|||||
80478 GGCTCTATCATGTCAGCCCTCATCAGCAGAGCGGCCACCATCAAGC 80527
|||||
430 lnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAla 446
|||||
80528 AGCTTCTCGCTTGTGGAGCTTCAAGTAAGATTGCTCCAGTGAAGCA 80577
|||||
447 ProAspAlaLysValArgMetValIleIleThrGlyProGluAlaG1 463
|||||
80578 CCAGATCTAAGTGAGGATGGTATGATCGCTGGATCAGCAGGGCTCG 80627
|||||
463 nPheLysAlaGlnGlyArgIleTyGlyLysIleLysGluGluAsnPheV 480
|||||
80628 GTTCAAGGCTCAGGAAGAT.TATGAAAAATCAAGAAGAAACTTCG 80676
|||||
480 alSerProLysGluGluValLysLeuGluAlaHisIleArgValProSer 496
|||||
80677 TTAGTCTCTAAGAAGAGGTGAAACTTGAAGCTCATATCAGATGCCATCC 80726
|||||
497 PheAlaAlaGlyArg.ValIleGlyLysGlyGlyLysThrValAsnGluL 513
|||||
80727 TTGCTCTCGCATTTACT...GAAAAGGAGGCAAAACCGTGAATGAAC 80773
|||||
513 euGlnAsnLeuSerSerAlaGluValValProArgAspGlnThrPro 529
|||||
80774 TTCAGATTTGTCAAGTCGAGAGTGTCTCTCCCTTGTGACCAGACACT 80823
|||||
530 AspGluAsnAspGlnValValLysIleThrGlyHisPheTyAlaCy 546
|||||
```

```
80824 GATGAGAATGACCAAGTGGTTGTCAAAATAACTGGTCACCTTCTATGCTT 80873
546 sGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnH 563
|||||
80874 CCAGTTGCCACAGAAAAATTCAGAAATTCCTCAGCTCAGTAAGCAGC 80923
|||||
563 isGlnGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
80924 AGCAACAACAGAAAGGCTCTGCAAGTGACCACTCAGTCAAGACGGAAG 80973
|||||
seq_name: gb_ov:AF064633
seq_documentation_block:
LOCUS AF064633 2697 bp mRNA VRT 03-JUN-1998
DEFINITION Xenopus laevis Vg1 RNA binding protein variant A mRNA, complete cds.
ACCESSION AF064633
VERSION AF064633.1 GI:3172446
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 2697)
AUTHORS Schwart, S., Standart, N.M. and Yisraeli, J.K.
TITLE RNA binding protein conserved in both microtubule and microfilament-based RNA localization
JOURNAL Genes Dev. (1998) In press
REFERENCE 2 (bases 1 to 2697)
AUTHORS Havin, L., Git, A., Elisha, Z., Oberman, F., Yaniv, K., Pressman
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) Biochemistry, Cambridge University, Tennis
COURT ROAD, CAMBRIDGE CB2 1GA, UK
FEATURES
source
1..2697
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/dev_stage="oocyte"
299..2080
/note="Vg1-RBP; Vg1 localization element binding protein"
/codon_start=1
/product="Vg1 RNA binding protein variant A"
/protein_id="AAC18597.1"
/db_xref="GI:3172447"
/translation="MNKLYIGNLSENVPDLESFLKESKIPFTQFLVKSGYAFVDC
PDETWMKAIDTSLGKVLHGKVIIEVHSVPKRQRSLQIRNIPPHQWELDSL
QYGTVENCEQVNTDSEAVVNVYANKHARQGLKNGYOLENYSLVKTYIPDENAT
POSQSOQLQPOOHQPGRRGFGPGAPARPGSAAARPKPOSEVRLMLVPTQFVGA
IGKEGATIRNITKOTQSKIDIRKENAGAAEKPTIHSTPEGCAACKIIMEIMQKE
AQDTFTTEIPLKILAHNNFVORLIGKERNLKIEQDPTDKITISPLQDLTYLPER
TITVKGSIETCAKEEENKKTRESYNDIAANLQAHILPGLNINLALGLFPSSSGM
PPPSAGVSPSTTSYSPPTQPESETVHLFIIPALAVGAILIKQGHKOLSRFAGS
IKIAPAEQPDALRMVITGPPEPAQKAGRIYGLKEENFPGKEEVEKLEAHIKVPS
YAAARGVIGKGTVELNQLTSAEVVPRDQTPDENQVVKITGHFYASLAQRKIQ
ETLAQVRRQQQQKTAQSGQPQPRK"
BASE COUNT 855 a 572 c 588 g 681 t 1 others
ORIGIN
alignment_scores:
Quality: 2483.00 Length: 593
Ratio: 4.515 Gaps: 6
Percent Similarity: 92.749 Percent Identity: 82.125
alignment_block:
US-09-685-696-176 x AF064633
Align seg 1/1 to: AF064633 from: 1 to: 2697
1 MetAsnLysLeuTyIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
```

|||||
299 ATGACACAGCTGTATATTTGGAAACCTTAGCGAAACAGTCAGGCCCCCGGA 348
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
349 TTTGGAAAGTCCTCTCAAGGAGTCGAAGATCCCTTTCCACGGGCGAGTTTC 398
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
399 TGGTCAAAAGTGTGACCGCTTTGTGGACTGTCCGGATGAGACGTGGGCT 448
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
449 ATGAAGGCATCGACACCTCTCAGGGAAGTGGAGTTGCATGGCAAGT 498
67 oLeuGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
499 GATACAAGTTGAACACTCAGTACCCCAAAAGGCAAGAGTCAAGAGCTTC 548
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
549 AGATAAGGAACATACCACTCATCTACAGTGGGAGGTACTGGACAGCCTA 598
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
599 TTAGCACAGTATGGACAGTGGAAACCTGTGAGCAAGTTAACACTGATTC 648
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
649 AGAAACTGCAGTAGTGATGTAACATATGCCAATAAGGAGCAGCGTAGAC 698
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPhetrLeuLys 150
|||||
699 AAGGACTCGAAAATTAACCGCTATCAGCTTGAGAACTATAGCCTGAAA 748
151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnG 167
|||||
749 GTGAGCTATACCTGATGAATGGCAACACCGCAATCACCATCCAGCA 798
167 n.....ProArgGlyArgArgGlyLeuGlyG 176
|||||
799 GCTTCAGCAGCGCAGCAGCAACACCTCAAGGGAGGCGTGGTTTGGGC 848
176 InArgGlySerSerArgGlnGlySerProGlySerValSerLysGlnLys 192
|||||
849 AAAGGGACCTGCCAGGAGGCGTCCCTGGTGTCTGCTGCAAGACCTAAG 898
193 Pro....CysAspLeuProLeuArgLeuLeuValProThrGlnPheValG 208
|||||
899 CCTCAGTCAGAGTTCCTCAGGATGCTGGTGGCCACACAGTTGTGTGG 948
208 yAlaIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnT 225
|||||
949 GCGGATCATTTGAAAGAGGGTGGCCACCATCAGGAACATCACTAAACAGA 998
225 hrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaGlu 241
|||||
999 CCCAGTCAAAAATAGATATTCACCGTAAAGAAAATGCTGGTGTGCAGAG 1048
242 LysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLy 258
|||||
1049 AAACCGATTAGATTTCATCAACCCCTCAAGGCTGTTTCAGCTGCATGCAA 1098
258 sSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrG 275
|||||
1099 GATCATTTATGAGATCATCGAGAGGAAGCTCAGGATACCAAGTTCACGT 1148
275 luGluIleProLeuLysIleLeuAlaHisAsnAsnPhetrValGlyArgLeu 291
|||||
1149 AAGAAATCCCTTAAAAATATTAGCACATAACAATTTGTTGGCGGTCTT 1198
292 IleGlyLysGluGlyArgAsnLeuLysIleGluGlnAspThrAspTh 308
|||||

seq_name: gb_ov:AF042353

seq_documentation_block:

1199 ATTGGGAAGAAGGAAGNACCTTAAGAAAATAGACCAAGACACACATAC 1248
308 rLysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluA 325
|||||
1249 CAAAATCAAAATATCTCCACTACAGGACTTGACACTGTACAATCTCTGAAC 1298
325 rGThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGluGlu 341
|||||
1299 GGACCATTTACAGTAAAGGCGACCATTTGACATGTGCCAAAGCCGAGAA 1348
342 GluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMe 358
|||||
1349 GAAGTCATGAAAAAATTAGGGAGTCCTATGAAAATGATATTGCTGCTAT 1398
358 tAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyL 375
|||||
1399 GAATCTACAGGCACATTTGATTCCTGGTTTAAACCTGAATGCTGGGCC 1448
375 euPheProPro....ThrSerGlyMetProProProThrSerGlyProPro 390
|||||
1449 TTTTCCACCATCTCTCGTCAGGAATGCCACCACCATCTGCTGGAGTTTCC 1498
391 SerAlaMetThr....ProProTyrProGlnPhe....GluGlnSerGluTh 405
|||||
1499 TCTCCGACAACATCTGCTTCTTATCCACATTTGGCGACGCCAGAGTC 1548
405 rGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleLeG 422
|||||
1549 AGAGACTGTTTCATCTCTTCATCCAGCTTTAGCTGTGGGAGCAATATTG 1598
422 lLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
|||||
1599 GAAAGCAAGGACACACATCAACAACACTGTCGCGTTTTCAGGAGCATCT 1648
439 IleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetValI 455
|||||
1649 ATTAAATTGCAACCGCAGAGGAGCTGTATGCCAACTAAAGAAATGGTGA 1698
455 eIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrG 472
|||||
1699 CATCACTGGACCTCTCTGAAGCCCAATTTAAGGCTCAAGGAAGGATCTATG 1748
472 lLysIleLysGluGluAsnPheValSerProLysGluGluValLysLeu 488
|||||
1749 GTAAACTTAAAGAGAAAACTTCTTTGGACCTTAAAGAGAGAGTGAACCTT 1798
489 GluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLy 505
|||||
1799 GAGGCCACATTTAAAGTCCGTCATATGCTGCTGGACGCTTATTGGCA 1848
505 sGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGluValV 522
|||||
1849 AGGAGCAAAACAGTAATGAATTAACAACTTAAAGTGCAGAAGTTG 1898
522 alValProArgAspGlnThrProAspGluAsnAspGlnValValLys 538
|||||
1899 TTGTGCCCGTGATCAAAACACCAAGATGAAATGATCAAGTGGTGTGTAAG 1948
539 IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnG 555
|||||
1949 ATAAGGGTCTACTTCTATGCTAGCCGCTTGCAAAAGGAAAAATTCAGGA 1998
555 uIleLeuThrGlnValLys.....GlnHisGlnGlnLysAlaLeuG 570
|||||
1999 AATACTGCTCAGGTAAAGAGACACAGCAGCAACAGCAGAGAGAGAGCGC 2048
570 InSerGlyProProGlnSerArgLys 579
|||||
2049 AAAGTGGACAAACCCCAACCAAGAGAAAA 2077

AF042353 3147 bp mRNA VRT 22-JAN-1998
 LOCUS xenopus laevis KH domain-containing transcription factor B3 mRNA,
 DEFINITION complete cds.
 ACCESSION AF042353
 VERSION AF042353.1 GI:2801765
 KEYWORDS African clawed frog.
 SOURCE Xenopus laevis
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 3147)
 Pfaff, S.L. and Taylor, W.L.
 AUTHORS Characterization of a Xenopus oocyte factor that binds to a
 TITLE developmentally regulated cis-element in the TFIIB gene
 JOURNAL Dev. Biol. 151 (1), 306-316 (1992)
 REFERENCE 92249652
 2 (bases 1 to 3147)
 Griffin, D. and Taylor, W.L.
 AUTHORS Direct Submission
 TITLE Submitted (11-JAN-1998) Biochemistry, University of Tennessee,
 JOURNAL Memphis, 858 Madison Ave., Memphis, TN 38163, USA
 FEATURES
 source
 1..3147
 /organism="Xenopus laevis"
 /product="KH domain-containing transcription factor B3"
 /db_xref="taxon:8355"
 /db_xref="GI:2801766"
 /cell_type="immature oocyte"
 972..2753
 /codon_start=1
 /product="KH domain-containing transcription factor B3"
 /protein_id="AAB97457.1"
 /db_xref="GI:2801766"
 /translation="MNKLYIGNLSNVSPDLESIFKESKIPFTGQFLVKSGYAFVDC
 PDEWAKAIDITLSGKVELHGKVIIEVSHVSKRORSKLIQIRNPPLHQLWELDSLILA
 QYGVNCEQVNTSEAVNVYTYANKEHARQLEKLNGLQENYSLKVTYIPDEMAT
 PQSFQLOQPQQHPQGRGFGORFARQSPGAAARPKQSEVPRLMLVPTQFVGA
 IIGKEGATIRIKQTSKIDIHRENAGAERKPTITHTPEGSAACKIIMEIQKE
 AQDTFTETIPKLIAHNHVFRLIGREGRNLKKIEQDTDKITISPLQDITLNPGR
 TITVKGSTETCAKAEVEMKKIRESYENDIAAMNLAHLIPLGLNLAALGLPPPSGGM
 PPSGASVSPPTTSASYPFGQOPESETVHLFI PALAVGAIIGKOGQHIKOLSRFAGS
 IKIAPAGQPAKLRLVITGPPEAQKAOGRHYCKLKEENFEFGPKESVLEAHLKVP
 YAGRVYKGGKTVNELNLNTSAEVVPRDTPDENQVVKITIGHFVASQLAQRKIQ
 EILQVRQQQQQKQKQSQGPQPRK"
 1607..1723
 /note="encodes KH domain 1"
 1847..2002
 /note="encodes KH domain 2"
 2243..2389
 /note="encodes KH domain 3"
 2489..2641
 /note="encodes KH domain 4"
 BASE COUNT 957 a 690 c 676 g 824 t
 ORIGIN
 misc_feature
 1607..1723
 /note="encodes KH domain 1"
 1847..2002
 /note="encodes KH domain 2"
 2243..2389
 /note="encodes KH domain 3"
 2489..2641
 /note="encodes KH domain 4"
 alignment_scores:
 Quality: 2483.00 Length: 593
 Ratio: 4.515 Gaps: 6
 Percent Similarity: 92.749 Percent Identity: 82.125
 alignment_block:
 US-09-685-696-176 x AF042353 ..
 Align seg 1/1 to: AF042353 from: 1 to: 3147
 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaProSerAs 17
 |||||
 972 ATGAAGAGCTGTATATTGGAACCTTAAGCAAGAGTGCAGCCCGCGGA 1021
 |||||
 17 pLeuGluSerIlePheIysAspAlaLysIleProValSerGlyProPheL 34
 |||||
 1022 TTGTGAAAGTCTCTCAAGAGTGAAGATCCCTTTCACCGGCGCAGTTTC 1071

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
 |||||
 1072 TGGTCAAAATGCGATACGCGTTTGTGGACTGTCGGGATGAGACGTGGCT 1121
 51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
 :|||
 1122 ATGAAGGCCATCGACACCCCTCTCAGGAAAGTGGAGTTGCATGGCAAGT 1171
 67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeu 84
 |||||
 1172 GATAGAAGTTGACACTCAGTACCCAAAGCAAGAGTTCGAAAGCTTC 1221
 84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
 |||||
 1222 AGATAAGGAACATACCCACCTCATCTACAGTGGGAGGTACTGGACAGCTA 1271
 101 LeuValGlnTyrClyValValGluSerCysGluGlnValAsnThrAspSe 117
 |||||
 1272 TTAGCACAGTATGGGACAGTGGAAACTGTGACGAAGTTAACACTGATTC 1321
 117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
 |||||
 1322 AGAACTGCGTAGTAGTGAATGTAACATATGCCAATAGGACGACGCTAGAC 1371
 134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
 |||||
 1372 AAGGACTCGAAAAATTAACGCGCTATCAGCTTGAGAACTATAGCCTGAAA 1421
 151 valAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
 |||||
 1422 GTGACGTATATACCTGATGAATGGCAACACCGCAATCACCATCCCGACGA 1471
 167 n.....ProArgGlyArgArgGlyLeuGlyG 176
 |||||
 1472 GCTTCAGACGCCCGCAGCAGCAACACCTCAAGGAGCGGTGGTGGGC 1521
 176 InArgGlySerArgGlnGlySerProGlySerValSerLysGlnLys 192
 |||||
 1522 AAAGGGACCTCGCAGCAGCGGGTCCCTGGTGTGCTGCTGCAAGACCTAAG 1571
 193 Pro...CysAspLeuProLeuArgLeuValProThrGlnPheValG 208
 |||||
 1572 CCTCAGTCAGAGGTTCACATGAGGATGCTGGTCCCCACACAGTTTGTGG 1621
 208 YAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnT 225
 |||||
 1622 GCGCATCATTTGGAAGAGAGGTGCCACCATCAGGAACATCCTAAACAGA 1671
 225 hrGlnSerLysIleAspValIleHisArgLysGluAsnAlaGlyAlaAlaGlu 241
 |||||
 1672 CCCAGTCAAAAATAGATATTTCACCGTAAAGAAAATGCTGGTGTGCAGAG 1721
 242 LysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLy 258
 |||||
 1722 AAACCGATTACGATTCACTCAACCCCTGGAAGGCTGTTTCAGTGCATCAA 1771
 258 sSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrG 275
 |||||
 1772 GATCATTTAGGACATCATGCAAGAGGAGCTCAGGATACCAAGTTCACTG 1821
 275 luGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeu 291
 |||||
 1822 AAGAAATCCCTTTAAAAATATTAGCACATAACAATTTTGTGGCCGCTTT 1871
 292 IleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspTh 308
 |||||
 1872 ATTGGGAAGAGGAAGAACCTTAAGAAAAATAGAGCAAGACACAGATAC 1921
 308 rLysIleThrIleSerProLeuGlnGluLeuLeuTyrLeuTyrAsnProGluA 325
 |||||
 1922 CAAATACAAATATCTCCACTAGGAGCTTGACACTGTACAATCCTGAAC 1971
 325 rgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGluGlu 341


```
|||||
1972 GGACATTACAGTAAAGGACGACATTGAGACATGTGCCAAAGCCGAGGAA 2021
|||||
342 GluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMe 358
|||||
2022 GAAGTCATCAAAAAAATTAGGAGTCCCTATGAAATGATATTGCTGCTAT 2071
|||||
358 tAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyL 375
|||||
2072 GAATCATAGGCACATTTGATTCCTGGTTTAAACCTGAATGCATTTGGGCC 2121
|||||
375 euPheProPro...ThrSerGlyMetProProThrSerGlyProPro 390
|||||
2122 TTTTCCCAACCTCTCGTCAGATGCCACCACTCTGCTGGAGTTCC 2171
|||||
391 SerAlaMetThr...ProProTyrProGlnPhe...GluGlnSerGluTh 405
|||||
2172 TCTCCGACAACATCTGCTTCTTATCCACCATTTGGGCAGCAGCCAGATC 2221
|||||
405 rGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleIleG 422
|||||
2222 AGAGACTCTTCACTCTTCACTCCAGCTTTAGCTGTGGAGCAATTATTTG 2271
|||||
422 lYlYsGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
|||||
2272 GAAGCAGGACACACATCAACCACTGTCGCTTTTGCAGGACATCT 2321
|||||
439 IleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetValIl 455
|||||
2322 ATTAAGATTGCACCGCAGAGGGACCTGATGCCAACTAAGATGGTGAT 2371
|||||
455 eIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrG 472
|||||
2372 CATCACTGGACCTCTCTGAAGCCCAATTTAAGGCTCAAGGAAGGATCTATG 2421
|||||
472 lYlYsIleLysGluGluAsnPheValSerProLysGluGluValLysLeu 488
|||||
2422 GTAACTTAAAGAGAAAGAACTTCTTTGGACCTAAAGAAAGTGAACCTT 2471
|||||
489 GluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLy 505
|||||
2472 GAGGCCACATTAAGTGCCTCATATGCTGCTGACGTGTTATTGGCAA 2521
|||||
505 sGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGluValV 522
|||||
2522 AGGAGGCAAAACAGATAATGAATCAAGATTTTAAACAAGTGCAGAAGTTG 2571
|||||
522 alValProArgAspGlnThrProAspGluAsnAspGlnValValValLys 538
|||||
2572 TTGTGCCCCGTGATCAACACCATGATGAAATGATCAAGTGGTTGTTAAG 2621
|||||
539 IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGl 555
|||||
2622 ATAACGGGTCACTTCTATGCTAGCCAGCTGCACAAAGGAAATTCAGGA 2671
|||||
555 uIleLeuThrGlnValLys.....GlnHisGlnGlnGlnLysAlaLeuG 570
|||||
2672 AATACTGGCTCAGTAAAGACAGACAGCAGCAGCAACAGCAGCAAGACGCG 2721
|||||
570 lNserGlyProProGlnSerArgArgLys 579
|||||
2722 AAAGTGGCAACCCCAACCAAGAAAGAAA 2750
```

seq_name: gb_ov:AF055923

seq_documentation_block: 2161 bp mRNA VRT 18-MAY-1998
LOCUS AF055923
DEFINITION Xenopus laevis RNA binding protein vera mRNA, complete cds.
ACCESSION AF055923
VERSION AF055923.1 GI:3136157
KEYWORDS African clawed frog.
ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.
1 (bases 1 to 2161)
Deshler,J.O., Highett,M.I., Abramson,T. and Schnapp,B.J.
A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates
Curr. Biol. 8 (9), 489-496 (1998)
JOURNAL
MEDLINE 98228351
REFERENCE 2 (bases 1 to 2161)
Deshler,J.O., Highett,M.I., Abramson,T. and Schnapp,B.J.
Direct Submission
TITLE Submitted (26-MAR-1998) Cell Biology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115, USA
JOURNAL
FEATURES
source Location/Qualifiers
1..2161
/organism="Xenopus laevis"
/db_xref="taxon:8355"
135..1919
/codon_start=1
/product="RNA binding protein Vera"
/protein_id="AAC41285.1"
/db_xref="GI:3136158"
/translation="MNKLYIGNLSNVSPDLESLEKSKIPETGQFLYKVGAFVDC PDETMAMKAIDTSLGKVELHGVIEVHSPVKRQRKLRQIRNIPHLQWEVLDSLLA QYGVENCQVNTSETAVNVNTYANKHAROGLEKNGYQLENYSKVTYIPDEMAT POAPSQLOQPOQHQPOGRGFGPARQSGAAARPKPQTEVPLRMLVPTQFVG AIIGKEGATIRNITKQSKIDIRKENAGAAEKPTIHTPEGCSAAKCIWEIMQK EAODTKFTFEEIPLKILAHNFVGRLLIGKSRNLKKIEODTDKTTISPLODLTYNPE RYITVKGSTIEPCAKAEEMMKIKRESYENDIAAMNLOAHLIPGLNINLALGFPSSSG MPSPSVSPSTSSYPPFQPOESETVHLFIPALAVAGAIIGKQOHKIQLSRFAGA SIKIAPAEQDAKLRMVIITGPPEAQFKAGRIYKGLKEENFFGPKVEEKLETHIKVP SYAAGRIGVGKGTVMELQNTSAEYVVPDQTPDENDEVVKITGHFVASQLAQRKI OEILAQVRROOQOQKQVQSGPOPPRRK"
BASE COUNT 710 a 477 c 491 g 483 t
ORIGIN
alignment_scores:
Quality: 2461.50 Length: 594
Ratio: 4.475 Gaps: 6
Percent Similarity: 92.593 Percent Identity: 81.313
alignment_block:
US-09-685-696-176 x AF055923 ..
Align seg 1/1 to: AF055923 from: 1 to: 2161
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
135 ATGAACAAGCTGTATATTGGAACCTCAGCGAAAACGCTGAGCCCAAGGA 184
|||||
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
185 TTTGGAAGTCTCTCAAGGAGTCTAAGATCCCTTTCACCGGCGAGTTTC 234
|||||
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
235 TGGTCAAAAGTGGATACGCGTTTCTGGACTCTCCGGATGACACGTGGCT 284
|||||
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
285 ATGAAGGCCATCGACACCTCTCAGGAAAAGTGGAGTTGTCATGCCAAAGT 334
|||||
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
335 GATTGAAGTCGAACACTCGGTACCCAAAAGCAAGGAGTGCAGAAAGCTTC 384
|||||
84 lNileArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
385 AGATACGAAACATACCGCCTCATCTACAGTGGGAGGTACTGGACAGCTG 434
|||||
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117

|||||.....
435 TTGCACAGTATGGCAGCTGAAAACCTGTGAGCAAGTGAACTGAATC 484

117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
485 AGAAACTGCCAGTACTGAATGTAACATATGCCAATAAGGACGCGTAGAC 534

134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
535 AAGGACTCGAAAAATTAAACGGCTATCAGCTGGAAAACTATAGCCCTGAAA 584

151 ValAlaTyrIleProAspGluMetAlaAlaGlnInAsnProLeuGlnGl 167
585 GTGACGTATATACCTGATGAATGGCAACACCGCAACGACCATCCCAAACA 634
167 n.....ProArgGlyArgArgGlyLeuG 175
635 GCTTCAGCAGCAGCCGAGCAGCAACACCTCAAGGAGCGCTGGGTTG 684

175 lyGlnArgLysSerArgGlnGlySerProGlySerValSerLysGln 191
685 GGCAAAAGGGGACCTGCCAGGACGGATCCCTGGTGTGCTGCAAGACCT 734

192 LysPro...CysAspLeuProLeuArgLeuLeuValProThrGlnPheVa 207
735 AAGCCTCAGACAGAGGTCCGCTGAGAATGCTGTGTTCCCACACAGTTTGT 784

207 lGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIlePheThrLysG 224
785 AGGGCGCATCATTCGGAAAGAGGGTGTACCATCAGAAACATCACTAAAC 834

224 lnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAla 240
835 AGACCCAGTCAAAAATAGATATTACCCGTAAGAAAAATGCAGGTGCTGCA 884

241 GluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCy 257
885 GAGAAACCGATTACAATTCACTCAACCCCTGAAGGCTGCTCAGCTGCATG 934

257 sLysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheT 274
935 TAAGATCATTTAGGATCATGCAGACAGGAAGCTCAGGATACCAAGTTCA 984

274 hrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 290
985 CTGAAGAAATCCCTTAAAAATATTAGCACATAACAACCTTTGTGGCCGT 1034

291 LeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluInaspThrAs 307
1035 CTTATTGGTAAGAAGGAAGGAACCTTAAGAAAAATTGAGCAAGATACAGA 1084

307 pThrLysIleThrIleSerProLeuGlnGluLeuThrLeuTyraSnProG 324
1085 TACAAAATACAATATCTCCACTACAGGACTTGACACTGTACAACTCCAG 1134

324 luArgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGlu 340
1135 AACGGACCATACAGTAAAGGCGACATTGAGCCATGTGCCAAGCCGAG 1184

341 GluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlase 357
1185 GAAGAAATCATGAAAAAATTAGGGAGTCTCTATGAAAATGATATTGCTGC 1234

357 rMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuG 374
1235 TATGAATTCAGGACCATTTGATTCTCTGTTTGAACCTGAATGCATTGG 1284

374 lyLeuPhePro...ProThrSerGlyMetProProThrSerGlyPro 389
1285 GCCTTTTCCCATCATCTCATCAGAAATGCCAACCACTTCTGTGGAGTT 1334

390 ProSer...AlaMetThrProProTyrProGlnPhe...GluGlnSerGl 404
|||||.....

| | | |
|------------------------------|---|------|
| 1335 | CCTCACCTACATCATCTACTTCTTTATCCACCATTGGGCGAGCAGCCAGA | 1384 |
| 404 | uThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleI | 421 |
| | ::::::::::::::::::::: | |
| 1385 | GTCAGAGACCGTTTCATCTCTTCATCCCAGCTTTAGCTGTGGGAGCGATTA | 1434 |
| 421 | leGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAla | 437 |
| | ::::::::::::::::::::: | |
| 1435 | TTTGGAAGCAAGGACAACACATCAAACAACATATCACGTTTTTCAGAGAGCA | 1484 |
| 438 | SerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetVa | 454 |
| | ::::::::::::::::::::: | |
| 1485 | TCTATTAAAGATTGCACCTCGGAGGAGCGCTGATGCCAACAATAAGATGGT | 1534 |
| 454 | IleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleI | 471 |
| | ::::::::::::::::::::: | |
| 1535 | GATCATCATCTGGACCTCCTGAAGCCCCAATTTAAGGCTCAAGGAAGGATCT | 1584 |
| 471 | yTGlyLysIleLysGluGluAsnPheValSerProLysGluGluValLys | 487 |
| | ::::::::::::::::::::: | |
| 1585 | ATGGTAAACTGAAGAAGAAAACCTTCTTCGGACCTAAAGAAGAAGTGAAA | 1634 |
| 488 | LeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGI | 504 |
| | ::::::::::::::::::::: | |
| 1635 | CTTGAGACCCCAATAAAGTGGCGTCTATGCTGCTGGACGGGTATTGG | 1684 |
| 504 | yLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGluV | 521 |
| | ::::::::::::::::::::: | |
| 1685 | CAAGGAGGAAAAACAGTAATGAACCTTCAGAAATTTAACAAAGTGCAGAGG | 1734 |
| 521 | aIvalValProArgaspGlnThrProAspGluAsnAspGlnValValVal | 537 |
| | ::::::::::::::::::::: | |
| 1735 | TTGTTGTGCCCGTGATCAAAACACCAGATGAAAATGATGAAGTGGTCGTT | 1784 |
| 538 | LysIleThrGlyHisPheTyraLaCySGlnValAlaGlnArgLysIleGI | 554 |
| | ::::::::::::::::::::: | |
| 1785 | AAGATAACGGGTCACTTCTATGCACTGACCTGTCACAAAGGAAAAATPCA | 1834 |
| 554 | nGluIleLeuThrGlnValLys.....GlnHisGlnGlnGlnLysAlaL | 569 |
| | ::::::::::::::::::::: | |
| 1835 | GGAAATATTGGCTCAGTGAAGAGACAGCAGCAGCAACACGACAGAAACAG | 1884 |
| 569 | eugLnSerGlyProProGlnSerArgArgLys 579 | |
| | ::::::::::::::::::::: | |
| 1885 | TGCAAAAGTGGACAACCCCAACCGAGAGAAAA 1916 | |
| seq_name: gb_ov:AF064634 | | |
| seq_documentation_block: | | |
| LOCUS | AF064634 2405 bp mRNA VRT 03 | |
| DEFINITION | Xenopus laevis Vg1 RNA binding protein variant D mRN | |
| KEYWORDS | cds, | |
| VERSION | AF064634 | |
| KEYWORDS | AF064634.1 GI:3172448 | |
| SOURCE | . | |
| ORGANISM | African clawed frog. | |
| | Xenopus laevis | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; | |
| | Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; | |
| | Xenopodinae; Xenopus. | |
| REFERENCE | 1 (bases 1 to 2405) | |
| AUTHORS | Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., P. | |
| | Schwartz,S., Standart,N.M. and Yisraeli,J.K. | |
| TITLE | RNA binding protein conserved in both microtubule an | |
| JOURNAL | microfilament-based RNA localization | |
| REFERENCE | Genes Dev. (1998) In press | |
| | 2 (bases 1 to 2405) | |
| AUTHORS | Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., P. | |
| | Schwartz,S., Standart,N.M. and Yisraeli,J.K. | |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (11-MAY-1998) Biochemistry, Cambridge Unive | |
| | Court Road, Cambridge CB2 1GA, UK | |
| FEATURES | Location/Qualifiers | |

```

source      1. .2405
/db_xref="Xenopus laevis"
/dev_stage="oocyte"
379. .2163
/notes="Vg1-RBP; Vg1 localization element binding protein"
/codon_start=1
/product="Vg1 RNA binding protein variant D"
/protein_id="AAC18598.1"
/db_xref="GI:3172449"
/translation="MNKLYIGNLSENVSPDLESIPKESKIPDTGFLVKSGVAFVDC
PDETAMKAIDTLSCVLEHGKVEHSVPKRQRSLQIRNIPHPLOWEYDLSLLA
QVGTVCNQVNTSTAVNNYANKHARQLEKINGIQLENYSLKVTYIPDENAT
PAQSOLQOQQHQPOGQFPGPARQSPGAAARPKQTEVPLRMLYPTQVQ
AIGKEGATIRNITQTKQKIDIRHKNAGAAEKPTIISTPEGCSAAACKIIMEIMQK
EQDITFEIPLKILAHNNFKIEGRNLKITEQDITKTIISPLQDLTLVNPQ
RTITVKSSTPECAKAEETMKIRESYENDIAMNLOAHLIPCLNLNALGLPSSSSG
MPPSVGVSPSPSSSYPPFGQOPESETVHLFIPALAVGAILGKQOQHILKQISRFAGA
SKIAPEAGDPAKLRVITITGPEAGQKQRIYGLKEENFFGPKKEEVKLETHIKVP
SYAAGRVIGKQKTVNELQNLTSAEVYVVRDQDTPDENDEVVVKITGHEFYASOLAQRKI
OEILAQVRQOQOQKTVSQGPQPRK"
BASE COUNT      763 a 537 c 537 g 568 t
ORIGIN

alignment_scores:
  Quality: 2461.50      Length: 594
  Ratio: 4.475          Gaps: 6
  Percent Similarity: 92.593      Percent Identity: 81.313

alignment_block:
US-09-685-696-176 x AF064634 ..
Align seg 1/1 to: AF064634 from: 1 to: 2405

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
379 ATGACAAAGCTGTATATTGGAAACCTACGCCAAACGTTGAGCCCCACGGA 428
|||||

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
429 TTTGGAAAGTCTCTTCAAGGAGTCTAAGATCCCTTTCACCGGGCAGTTTC 478
|||||

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
479 TGGTCAAAAGTGGATACGCGTTTGGCACTGTCCGGATGAGACGTGGGCT 528
|||||

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
529 ATGAAGGCCATCGACACCCTCTCAGGGAAGTGGAGTTGCATGCCAAAGT 578
|||||

67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
579 GATTGAAGTCGAACACTCGGTACCCAAAGGCAAGGAGTCGAAAGCTTC 628
|||||

84 InIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
629 AGATACGAAACATACCGCCTCATACAGTGGGAGGTACTGGACAGCGCTG 678
|||||

101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
679 TTGGCAGATGAGCAGTGGGAACCTGTGAGCAAGTGAACACTGAATC 728
|||||

117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
729 AGAAATGCACTAGTGAATGTAACATATGCCAATAAGGAGCACCGCTAGAC 778
|||||

134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
779 AAGGACTCGAAAAATTAAACCGCTTACAGCTGGAAAAATATAGCGCTGAAA 828
|||||

151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
|||||

```

```

829 GTGACGTATATACCTGTGTAATGCCAACACCGCAAGCACCATCCCAACA 878
167 n.....ProArgGlyArgArgGlyLeuG 175
      |||:|||||
879 GCTTCAGCAGCAGCGCAGCAGCAACACCCCTCAAGGAGCGTGGGTTTG 928
175 lyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysGln 191
      |||:|||||
929 GGCAAAAGGGGACCTGCCAGGAGGATCCCTGCTGCTGCTGCAAGAGCT 978
192 LysPro...CysAspLeuProLeuArgLeuLeuValProThrGlnPheVa 207
      |||:|||||
979 AAGCCTCAGACAGAGGTTCCCTGAGAAATGCTGGTCCCAACACAGTTTGT 1028
207 lGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
      |||:|||||
1029 AGGGCGCATCATTTGGGAAAGGGTGTCTACCATCAGGAACATCACTAAC 1078
224 InThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAla 240
      |||:|||||
1079 AGACCCAGTCAAAAATAGATATATCACCGTAAAGAAATGCAGGTGCTGCA 1128
241 GluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCy 257
      |||:|||||
1129 GAGAAACCGATTACAATTCACCTCAACCCCTGAAGGCTGCTCAGCTGCATG 1178
257 sLysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheT 274
      |||:|||||
1179 TAAGATCATTTATGGAGATCATGCAGAGGAAGCTCAGGATACCAAGTTCA 1228
274 hrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 290
      |||:|||||
1229 CTGAAGAAATCCCTTAAAAATATTAGCACATAACAACTTTGTTGGCGGT 1278
291 LeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAs 307
      |||:|||||
1279 CTTATTTGGTAAAGAAAGGAAGAACCTTTAAGAAATTTGAGCAAGATACAGA 1328
307 pThrLysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProG 324
      |||:|||||
1329 TACAAAATCACAATATCTCCACTACAGGACTTGACACTGTACAAATCCAG 1378
324 luArgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGlu 340
      |||:|||||
1379 AACGGACCATTTACAGTAAAGGAGCAGCATTTGAGCCATGTGCCAAAGCCGAG 1428
341 GluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlase 357
      |||:|||||
1429 GAAGAAATCATGAAAAAATTTAGGAGGTCCTATGAAAAATGATATTGCTGC 1478
357 rMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuG 374
      |||:|||||
1479 TATGAATCTCCAGGCACATTTGATTCTCTGTTTGAACCTGAATGCATGG 1528
374 lyLeuPhePro...ProThrSerGlyMetProProThrProThrSerGlyPro 389
      |||:|||||
1529 GCCTTTTCCCATCATCTCATCAGGAATGCCACCACCTTCTGTTGGAGTT 1578
390 ProSer...AlaMetThrProThrProGlnPhe...GluGlnSerGl 404
      |||:|||||
1579 CCCTCACCTACCTCATCTACTCTTATCCACCANTTTGGGAGCAGCCAGA 1628
404 uThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleI 421
      |||:|||||
1629 GTCAGAGACCGCTTCATCTCTCATCCCAAGCTTTAGCTGTGGAGCGATTA 1678
421 leGlyLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGlyAla 437
      |||:|||||
1679 TTGAAAGCAAGGACACACATCAAAACAACTATCACGTTTTCAGGAGCA 1728
438 SerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetVa 454
      |||:|||||
1729 TCTATTAGATTGCACTGCGGAGGAGCTGATGCCAAACTAAGAAATGGT 1778

```

454 l l l e l l e t h r g l y p r o g l u a l a g l n p h e l y s a l a g l n g l y a r g i l e t 471
|||||
1779 g a t c a t c a c t g g a c c t c c t g a a g c c a a t t a a g g c t c a a g a a g g a t c t 1828
|||||
471 y t g l y l y s l e l y s g l u g l u a s n p h e v a l s e r p r o l y s g l u g l u v a l l y s 487
|||||
1829 a t g g t a a c t g a a g a a g a a a c t t c t c g g a c c t a a a g a a g a a g t g a a 1878
|||||
488 l e u g l u a l a h i s l e a r g v a l p r o s e r p h e a l a a l a g l y a r g v a l l e g l 504
|||||
1879 c i t t g a c a c c a c a t t a a a g t g c c g t c a t a t g t g c t g a c g g t t a t t g g 1928
|||||
504 y l y s g l y l y s t h r v a l a s n g l u l e u g l n a s n l e u s e r s e r a l a g l u v 521
|||||
1929 c a a g a g a g a a a a c a g t a a t g a a c t t c a g a a t t t a a c a a g t g c a g a g 1978
|||||
521 a l v a l v a l p r o a r g a s p g l n t h r p r o a s p g l u a s n a s p g l n v a l v a l 537
|||||
1979 t t g t t g t g c c c g t g a t c a a a c a c c a g a t g a a a t g a t a a g t g g t c g t t 2028
|||||
538 l y s i l e t h r g l y h i s p h e t y r a l a c y s g l n v a l a l a g l n a r g l y s i l e g l 554
|||||
2029 a g a t a a c g g t c a c t t c t a t g c a a g t c a g t t g c a a a g a a a a t t c a 2078
|||||
554 n g l u l l e u t h r g l n v a l l y s g l n h i s g l n g l n l y s a l a l 569
|||||
2079 g g a a a t a t t g g c t c a g t a a g a a g a c a g c a g c a c a c a g a a a c a g 2128
|||||
569 e u g l n s e r g l y p r o g l n s e r a r g a r g l y s 579
|||||
2129 t g c a a a g t g g a c a a c c c a c c a c c a g a a g a a a 2160
|||||

seq_name: gb_ov:AF161270

seq_documentation_block: 2011 bp mRNA VRT 27-JUL-1999
LOCUS AF161270
DEFINITION Danio rerio Vgl RNA binding protein mRNA, complete cds.
ACCESSION AF161270
VERSION AF161270.1 GI:5596631
KEYWORDS
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
AUTHORS Zhang, Q., Yaniv, K., Oberman, F., Wolke, U., Git, A., Fromer, M.,
Taylor, W., Meyer, D., Standart, N., Raz, E. and Yisraeli, J. K.
TITLE Vgl RBP intracellular distribution and evolutionarily conserved
expression suggest multiple roles during development
Mech. Dev. (1999) In press
REFERENCE
AUTHORS Oberman, F., Fromer, M. and Yisraeli, J. K.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1999) Anatomy and Cell Biology, Hebrew University
Hadassah Medical School, POB 12272, Jerusalem 91120, Israel
FEATURES
source
1. .2011
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="linkage group 19"
/map="between Z3023 and Z15451"
1. .1749
/note="Vgl RBP"
/codon_start=1
/product="Vgl RNA binding protein"
/protein_id="A045610.1"
/db_xref="GI:5596632"
/translation="MNKLYIGNVSEQAALDLESIFQWKIPFSAPFLVKSQAFVDC
PDEKVARADITLGSVKGLVKLEVEHSFKRSCKLQIRNIPPHQWELVDGLIA
QYGVSECEQVNTDETAVVNVRYGAKDQAREAMDKLNGFLMENYALKVSIYPDETAA
ADAPAVGRRGRGNRPGRGSPSLGARPKLQSDVPLRLLLVPTQFVGAIIIGKEGATIR

NITKQTHSKIDIHRKENAGAAEKPIVHSTPEGCSSACRNIMEIMQKRAIDTKITEEI
PLKILAHNNFVRLIGEGRNKKIEQDQTDKTIITSPLODITLYNPERTITVKGTLDA
CAKAEETIMKKVRESYENDVAAMHLQSNLIPGLNLALEPFGAASGSPSVSGPP
PQAQGYOSFGAOMESYTHLFI PALAVGAIIGKOGHILKLSRFAGASIKIAPADGI
DAKQMWIISGPEAPKACRIFGKLKEENFGPKKEVKLEAHIKVPSEARVIGK
GGKTVNELONLTSAEVVVPRDQTPDENDQVVKITGHFYASQLAKRQKQEIISQVRRQ
QQPRPSAAGPPVARRK"

BASE COUNT 581 a 494 c 501 g 435 t
ORIGIN

alignment_scores:
Quality: 2261.50 Length: 584
Ratio: 4.299 Gaps: 5
Percent Similarity: 90.068 Percent Identity: 77.055
alignment_block:
US-09-685-696-176 x AF161270 ..
Align seg 1/1 to: AF161270 from: 1 to: 2011

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
1 ATCAATAAGCTGTACATCGGGAACGTGAGCGAGCGGCGCGCTGGA 50
|||||
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
51 CTTGGAAAGTATCTTTGAGCAGCTGGGAATACCTTTTCAGTGTCCCTCC 100
|||||
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
101 TCGTAAAAAGTGGCTACGGCTTTGTGGATTGCCCGCAGCAAGAGGTAGC 150
|||||
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
151 ATGAGGGCCATGACACGCTTTTCAGGAAAAGTAGAACTCCACGGAAAAAGT 200
|||||
67 olleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
201 CTTAGAAGTGGAGCAGCTCTGCTTAAGCACAAGAGGAGCTGTAAAGCTG 250
|||||
84 lnLeArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
251 AGATCAGGAATATCCACCTCATCATGAGTGGGAAGTTTATAGCGGTTG 300
|||||
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
301 TTAGCACAGTATGGCAGCTGGAAGCTGTGAACAAGTCAACACTGCACAC 350
|||||
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
351 CGAGACAGCAGCTGCTCAATGTGGCATATGGTCCCAAGGACCAAGCCAGAG 400
|||||
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
401 AAGCAATGACAAATTAATATGGCTTCCTTATGGAGAATACAGCAGCTAAAG 450
|||||
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
|||||
451 GTTTCCTATATCCAGATGAGCTGTCGAGCTGTATGCCCT.....GC 494
|||||
167 nProArgGlyArgArgGlyGlnArgGlySerSerArgGlnGlyLys 184
|||||
495 TGTGGGTGGCAGGAGGCTTTAAACCCAGCTGGGCCACCTCGCCCAAGGCT 544
|||||
184 erProGlySerValSerLysGlnLys....ProCysAspLeuProLeuArg 199
|||||
545 CTCCCAGTCTGGCGCTAGACCTAAACTACAGTCAGACGTCGCTCTACGC 594
|||||
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
|||||
595 TTGCTGGTACCCACCGAGTTGTAGGAGCTATTATTGGCAAGGAAGGTGC 644
|||||

```
216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
|||||
645 TACCATCGCAACATCAACCAAGCAGCACCCACTCAAAAGATTGACATTTCATC 694
|||||
233 rgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
|||||
695 GTAAGGAGAACGACAGGACGACGAGAGCAACATCACTGTTCACTCAACA 744
|||||
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLy 266
|||||
745 CCAGAGGCGCTCGATTGCTGCTGCGAAACATCATGAGATCATGCAGAA 794
|||||
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
|||||
795 AGAGCAATCGACACAAAGATCACTGAAGAAATTCCTCAAAATTTTGG 844
|||||
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
|||||
845 CTCATAACAATTTGTTGGCGCCTCATTTGAAAGAGGAGCGCAACTTG 894
|||||
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG1 316
|||||
895 AAGAAATCGAACAAGATCTGACACCAAGATCACATCTCACCTCTGCA 944
|||||
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
|||||
945 GGATCTGACCTTGTATATACCCAGAGCGCACCATTAATCTGTTAAAGAACGT 994
|||||
333 alGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArgGlu 349
|||||
995 TAGATGCTGTGCAAAAGCTCAAGAAAGAAATCATGAAGAAAGTCCGCGAG 1044
|||||
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
|||||
1045 TCCTATGAGAACGATGTAGCTGCCATGCACCTTCAATCTAACCTGATTCC 1094
|||||
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhePro...ProThrSerGly. 381
|||||
1095 AGGGCTCAATCTGAATGATGGGTTTATTTCTGTCGACGAGTGGAG 1144
|||||
382 .....MetProProThrSerGlyProProSerAlaMetThrProPro 396
|||||
1145 GGATTTACCTCTGTGTATACGACACACCGCTGGGGCTCAAGCTGGA 1194
|||||
397 TyrProGlnPhe...GluGlnSerGluThrGluThrValHisGlnPheI1 412
|||||
1195 TATCAATCATTTGGGGCCACATGAGTGCAGAACTGTGACCTCTTCAT 1244
|||||
412 eProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleL 429
|||||
1245 TCCTGCACTGGCAGTTGGTGCTATTATCGCAAGCAGGGCCCAACACATTA 1294
|||||
429 yGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlu 445
|||||
1295 AACAGCTCAGCGCTGCTGCTGGTGCATCAAAATTCGCCCTGCTGAT 1344
|||||
446 AlaProAspAlaLysValArgMetValIleIleThrGlyProGluAla 462
|||||
1345 GGAATTGATGCCAAGCAAGAAATGGTATTATTCTGGACACACAGAGGC 1394
|||||
462 agLnpHelysAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnP 479
|||||
1395 ACAGTTTAAGGCTCAGGACGATATCTTTGGAAGACTTAAAGAGAGAACT 1444
|||||
479 heValSerProLysGluGluValLysLysLeuAlaHisIleArgValPro 495
|||||
1445 TCCTCGGACCAAGAGAGGATTAATAGAGGCTCATATAAAGTGCACA 1494
|||||
496 SerPheAlaLaGlyArgValIleGlyLysGlyGlyLysThrValAsnG1 512
|||||
1495 TCATTCGCTGCTGCGACAGTCTATTGGTAAGAGGACCAAAACGGTGAACGA 1544
|||||
512 uLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnThrP 529
```

```
|||||
1545 ACTACAGAACTTGACCTGCGGAGTGGTCCCGGAGGACACAGAC 1594
|||||
529 roAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyrAla 545
|||||
1595 CTGATGAAAATGACCAAGTAGTAGTAAATAATCACAGGCCACTTCTATGCA 1644
|||||
546 CysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysG1 562
|||||
1645 AGCCAGCTGGCCCGAGAGAAATCCAGGAGATTATCTCTCAGGTGAGCGG 1694
|||||
562 nHisGlnGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgArgL 579
|||||
1695 GCAACAGCAGCCCAAGCCCTCAGCGGAGGTCCACCGAGTCCGCCAGGAGGA 1744
|||||
579 ys 579
||
1745 AA 1746

seq_name: gb_ov:AF026527
seq_documentation_block:
LOCUS AF026527 2021 bp mRNA VRT 30-OCT-1997
DEFINITION Gallus gallus zipcode-binding protein (ZBP1) mRNA, complete cds.
ACCESSION AF026527
VERSION AF026527.1 GI:2570920
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2021)
AUTHORS Ross A.F., Oleynikov, Y.S., Kislauskis, E.H., Taneja, K.L. and
Singer, R.H.
TITLE Characterization of a beta-actin mRNA zipcode-binding protein
JOURNAL Mol. Cell. Biol. 17 (4), 2158-2165 (1997)
MEDLINE 9722007
REFERENCE 2 (bases 1 to 2021)
AUTHORS Oleynikov, Y.S., Ross, A.F. and Singer, R.H.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1997) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
source 1..2021
/organism="Gallus gallus"
/db_xref="taxon:9031"
/cell_type="fibroblast"
/dev_stage="embryo"
1..2021
/gene="ZBP1"
129..1859
/gene="ZBP1"
/function="binds a zipcode element in beta-actin mRNA
3'UTR and is involved in its localization"
/notes="beta-actin mRNA zipcode-binding protein"
/codon_start=1
/product="zipcode-binding protein"
/protein_id="AAB82295.1"
/db_xref="GI:2570921"
/translaton="MNKLYIGNLNESTPADLEKVFNDHKISFGQFLVKSQYAFVDC
PDEQWAMKAIEFTSGKVELHGOLEIETHSVPKRSRKIQIRNIPQLRWEVLDGLLA
OYGTWENECOVNTDSEYAVNVYTNREOTROAIMKLNHOLENHVKVSYIPDEQSV
QCPENGRGCGFARGAPROGSPVTAGAPKQPDPLRLVPTOVVGVGAIKGEATI
RNITKOTQSKIDVHRKENAGAAKAIHSIHSTPECCSNACKMILEIMQKAKDKPTADE
VPLKILAHNNFVGRLLIGRNLKRVQDTEKRTITISSLODUTLYINPERTITVKGSTIE
NCCKAEQIEIMKKVREAYENDVAAMSQLSHLIPGLNLAAYGLFPASSNAVPPPPSSVSG
AAPYSFSPPEQETVHFVIPAQAVGAIIGKGOHIGKLSRFASASIKIAPPETPSKV
RMVYITGPPEAQFKAQGRIGYGLKLEENFFGPEEVLKETHIRVPSAAGRVIGKGKGT
VNELQNLTAEEVVVPRDQTPDENEQVIVLIIGHFYASQMAQRKIRDILIAQVKQHQHG
QSGQLQARRK"
138..155
/misc_feature
/gene="ZBP1"
```

```

misc_feature      /note="encodes RNP2 RNA binding domain"
236..260
/feature="ZBP1"
misc_feature      /note="encodes RNP1 RNA binding domain"
354..374
/feature="ZBP1"
misc_feature      /note="encodes nuclear localization signal; NLS"
621..632
/feature="ZBP1"
misc_feature      /note="encodes nuclear localization signal; NLS"
747..773
/feature="ZBP1"
misc_feature      /note="encodes KH-domain #1"
990..1016
/feature="ZBP1"
misc_feature      /note="encodes KH-domain #2"
1062..1088
/feature="ZBP1"
misc_feature      /note="encodes REV-like nuclear export signal; NES"
1374..1400
/feature="ZBP1"
misc_feature      /note="encodes KH-domain #3"
1620..1646
/feature="ZBP1"
BASE COUNT      525 a 562 c 583 g 351 t
ORIGIN

```

alignment_scores:

```

Quality: 2219.50      Length: 584
Ratio: 4.196          Gaps: 7
Percent Similarity: 90.582 Percent Identity: 74.829

```

alignment_block:

```
US-09-685-696-176 x AF026527 ..
```

```
Align seg 1/1 to: AF026527 from: 1 to: 2021
```

```

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
129 ATGAACAACTGTACATCGGACCTTAAACGAGAGCGTGACTCGGCCGA 178
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
179 CTGGAGAAAGTTTCAACGACCACAAGATCTCCTTCAGCGGGCAGTTC 228
34 euVallyThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
229 TGGTCAAGTCCGGCTATGCTTTTGTGGACTGCCCGCAGCAGTGGGCC 278
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
279 ATGAAGCCATCGAACTTTTCGGGGAAGTGGAGCTGCATGGAACA 328
67 orLeuGluValGluHisSerValProLysArgGlnArgIleArgLysLeu 84
329 GCTTGAGATTGAACACTCAGTGCTTAAAGCAAGAGCCGGAATTC 378
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
379 AAATCCGAATAATCCACCCAGCTCGCATGGAGGTTCTGGATGGCTTG 428
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
429 CTAGCAATATGGCACTGTAGAAAATCTGGACAAGTGAACACAGACAG 478
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
479 TGAGACCGCTGTGGTTAATGTCACCTACACAACCGGGAGCAGACCGGC 528
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150

```

```

529 AAGCCATCATGAAGTTAAACGGGCACACAGCTGGAGAACACAGTGTGAAAA 578
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
579 GTCTCTACATCCTCATGAGCACTCCGTG...CAGGCCACAGAAATGG 625
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly 184
626 CGCGCGGGGT...GGCTTTGGGCGCCGAGGTGCCCCCGCAGGGCT 669
184 erPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
670 CCCCTGTCACTGCAGGGGTCCAGTCAAGCAGCAGCCGCGTGGACATCCG 719
198 LeuArgLeuValProThrGlnPheValGlyAlaIleIleGlyLysG 214
720 CTCCTGCTGCTGGTGGCCAGCAGTATGTGGTGCCATCATTTGGCAAGGA 769
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
770 AGGGGCCACCATCAGGAACATTACCAAGCAGACACAGTCCAAGATTGATG 819
231 aHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
820 TGCACCGTAAAGGAATGCAGGAGCTGCAGAAAAGCTATCAGCATCCAC 869
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 264
870 TCCACCCCGAGGGCTGCTCTGCTGCTGCAAAATGATCTTGGAGATCAT 919
264 tHisLysGluAlaGlnAspIlePheThrGluGluIleProLeuLysI 281
920 GCAGAAGGAGGCAAGACAGACAGTGTGATGAAGTGCCTCTGAAAA 969
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
970 TCTTGGCCCATAAACATTTTGTGGGGCGCTGATTGGCAAGAAGGCGG 1019
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
1020 AACTTGAAGAAAGTGGAGCAGGATACTGAGACAAAATCACCATCTCATC 1069
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
1070 CTTGCAAGGACCTGACCTGTACAAACCCGAAAGACGATCACAGTGAAG 1119
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
1120 GCTCTATCGAGAACTGCTCAAAAGCAGCAGGAGATCATGAAGAAGTG 1169
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHis 364
1170 AGAAGAGCTTACGAGAACGATGTTGAGCCATGAGCCATGCAATCTCATCT 1219
364 ulleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerG 381
1220 CATCCCTGGCTCAACCTGGCTGCAAGTGGCCCTTCTCCCTGCTCTCCA 1269
381 lyMetProProProThrSerGlyProProSerAlaMetThr.....Pro 395
1270 ATGACATGACCTCT.....CTCCCAAGCAGCGTCTCCGGGGCTGCT 1310
396 ProTyrProGlnPheGluGlnSerGluThrGluThrValHisGlnPheI 412
1311 CCATACAGCTCTTTCATGCTCCGGAGCAGGAGCCGTCACAGCTCTTCAT 1360
412 eProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlnHisIle 429
1361 CCCTGCCCAAGCGGTGGTGGCCATCATTTGGCAAGAGGGCCAGCACATCA 1410
429 yGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlu 445
1411 AGCAGCTCTCCCGTTTCGCAAGTGCCTCTATTAAAGATTGCACCCCGGAG 1460

```



```
|||||
381 AGATCCGCAATATTCACCTTCAGCTCCGATGGAAGTGTAGATAGCCTG 430
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
431 CTGGCTCAGTACGTCAGTGGAGAACTGTGACAAAGTGAACACTGAAAG 480
117 rGluThrAlaValAlaValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
481 TGAGACAGCGGTGGTCAACGTCACCTACTCTAACCAGGAGCACCAGGC 530
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
531 AAGCTATCATGAAGCTAAATGGCCATCACTGGAGAAACCATGCCCTGAAG 580
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
581 GTCTCTACATACCTGATGACGAGATAACA...CAAGGTCTCTGAGAAATG 627
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly 184
|||||
628 GCCTCGTGA...GGCTTTGGTCTCGGGGCCAGCCCGGCAAGGT 671
184 erPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
|||||
672 CGCCCGTGGCAGCAGGGGTCTCCAGCCAAAGCAGCAGCGAGTGGACATCC 721
198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysG 214
|||||
722 CTCGGCTCTGTGTGCTACGCAAGTATGTAGCGCTATCATTTGGCAAGGA 771
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
|||||
772 GGGTGCACCATCCGAACATCACAAACAGACAGCGAGTCCANATAGACG 821
231 aHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
|||||
822 TGCATAGGAAGGAATGGCGGCGCTCGGAGAAAGGCCATCAGCGTGCAT 871
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 264
|||||
872 TCAACCCCTGAAGCGTCTCTCCGCGTCAAGATGATCTTGGAGATTAT 921
264 tHisLysGluAlaGlnAspLysPheThrGluGluIleProLeuLysI 281
|||||
922 GCACAAAGGAGCAAGACACCAACAGCGCAGATGAATTCCTCCCTGAAGA 971
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
|||||
972 TCCTGGCTCATAAACATTCTGTCGGCGACTCATTTGGCAAGGAGCGCG 1021
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
|||||
1022 AACCTGAAGAGGTGGAGCAGGACAGACAGACAGACCAAGATCACCATCTC 1071
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
|||||
1072 GCTCCAGGACCTCACGCTCTATAACCTTGAGAGACCACATCCTGTGAAG 1121
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||
1122 GCGCCATTGAGAACTGTTCAGGGCGGAGCAGGAGATCATGAAGAAGTT 1171
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
|||||
1172 CGAGAGGCTTACCAGAACGACGTGGCGCCATCAGCTTCAGTCCACCT 1221
364 uIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerG 381
|||||
1222 CATCCCTGGGCTTAACCTGGCTGCTGTAGTCTCTCCAGCTTCATCCA 1271
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
:: |||||
```

```
1272 GCCTGTCTCCCTCT.....CTCCCCAGCAGTGTCACTGGGGCTGCT 1312
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
|||||
1313 CCCTATAGCTCCTTCATGCAAGCTCCGAGAGTGGTATGCAAGTGT 1362
411 eIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisI 428
|||||
1363 CATCCCCGCCAGGCTGTGGCGCCATCATTTGCAAGAGGCGCAGCACA 1412
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
|||||
1413 TCAACAACACTCTCCCGTTCCGAGCGCTCCATCAAGATTGCTCCACCA 1462
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProG 461
|||||
1463 GAACACCTGACTCCAAAGTTCGAATGCTCATCTACCTGACCCCCAGA 1512
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGlu 478
|||||
1513 GGCTCAGTTCAAGGCTCAGGGAAGATTTATGCAAACTAAAGAAGAGA 1562
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
|||||
1563 ATTCTCTTTGGTCCCAAGGAGAAAGTAAAGCTAGAGACCCACATACGG 1612
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValas 511
|||||
1613 CCGGCTTCAGCAGCGCGGCGCTCATCGGCAAGGCGCAAAACGGTGAA 1662
511 nGluLeuGlnAsnLeuSerSerAlaGluValValValProArgAspGlnT 528
|||||
1663 TGAGCTGCAGAACTTGACTGCAGCTGAGGTGGTAGTGCCAAAGAGACC 1712
528 hrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyr 544
|||||
1713 CCGCGATGAGAAACGACCAAGTCAATTTAAGATCATCGGACATTTTAT 1762
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLy 561
|||||
1763 GCAGCGCAGATGGCTCAGCGGAAGATCCGAGACATCTCTGCTCAAGTTAA 1812
561 sGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgA 578
|||||
1813 GCAA...CAGCACCAGAAGGA...CAGACCAACTGGCCCGCAGCTCGGA 1856
578 rGlys 579
|||||
1857 GGAAG 1861

seq_name: gb_pr:AF117106
seq_documentation_block: 2130 bp mRNA PRI 26-JAN-1999
LOCUS AF117106
DEFINITION Homo sapiens IGF-II mRNA-binding protein 1 (IMP-1) mRNA, complete cds.
ACCESSION AF117106
VERSION AF117106.1 GI:4191607
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2130)
AUTHORS Nielsen,J., Christensen,J., Lykke-Andersen,J., Johnsen,A.H.,
Wewer,U.M. and Nielsen,F.C.
TITLE A family of insulin-like growth factor II mRNA-binding proteins
represses translation in late development
JOURNAL Mol. Cell. Biol. 19 (2), 1262-1270 (1999)
MEDLINE 99108099
REFERENCE
2 (bases 1 to 2130)
AUTHORS Nielsen,J., Christensen,J., Lykke-Andersen,J., Johnsen,A.H.,
Wewer,U.M. and Nielsen,F.C.
```



```
|||||
1292 TCAACAGCTCCCGGTTTGGCAGCGCTCCATCAAGATTCACACACC 1341
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProG 461
|||||
1342 GAAACACCTGACTCCAAAGTTCGTATGTTATCATCTACCTGGACCGCAGA 1391
461 uAlaGlnPheLysAlaGlnGlyArgIleThrGlyLysIleLysGluLys 478
|||||
1392 GGCCCAATCAAGGCTCAGGAAGAATCTATGCAAACTCAAGGAGGAGA 1441
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
|||||
1442 ACTTCTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACCTGTG 1491
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAs 511
|||||
1492 CCAGCATCAGCAGCTGCGCGGTTCATGCAAGGTGGAAAACGGTGAA 1541
511 nGluLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnT 528
|||||
1542 CGAGTTCAGAAATTCACGCGCAGCTCAGGTGGTAGTACCAAGAGACCAGA 1591
528 hrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheThr 544
|||||
1592 CCCTCATGAGAACGACGACGATCATGCTGAAATCATCGACATTTCTAT 1641
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnVal 561
|||||
1642 GCAGTCAGATGCTCAACGGAAGATCCGAGACATCTCGCCAGGTTAA 1691
561 sGlnHisGlnGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArg 578
|||||
1692 GCAG...CAGCATCAGAAGGA...CAGAGTACACGCGCCACGACGGA 1735
578 rGlys 579
|||||
1736 GGAAG 1740
```

seq_name: gb_pr:AF198254

```
seq_documentation_block: 2381 bp mRNA PRI 02-MAR-2000
LOCUS AF198254 Homo sapiens mRNA-binding protein CRDBP mRNA, complete cds.
DEFINITION AF198254
ACCESSION AF198254
VERSION AF198254.1 GI:7141071
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2381)
Ioannidis,P., Trangas,T., Dimitriadis,E., Samiotaki,M.,
Panoutsakopoulos,G., Kyriazoglou,I., Voutzoullas,S.,
Tsilapalis,C.M., Kittas,C., Agnantis,N. and Pandis,N.
Ecotopic expression of a KH-domain containing protein, highly
homologous to both human IMP-1 and mouse CRD-BP, in benign and
malignant mesenchymal tumors
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2381)
AUTHORS Ioannidis,P., Trangas,T., Dimitriadis,E. and Samiotaki,M.
TITLE Direct Submision
JOURNAL Submitted (25-OCT-1999) G. Papanikolaou Research Center, St Savvas
Hospital, 171 Alexandras Avenue, Athens, Attica 11522, Greece
FEATURES
source
1..2381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21.3-q22"
/cell_line="MCF-7"
306..2039
/note="KH-domain containing protein"
```

```
/codon_start=1
/product="mRNA-binding protein CRDBP"
/protein_id="AAF37203.1"
/db_xref="GI:7141072"
/translation="MNKLYIGNLNEVTPADLEKVFABHKISYSGOFLVKSGYAFVDC
PDERHMKAIETFSGKVELQKRLIEHSVPKQKORSRKIQIRNPPQLRWELVDSLLA
OYGTVENCEQVNTSETAVNVVTSNREQTROAIMKLNHOLENHALKVSYPDEQIA
QPGNRRGGFGSGRQSPVAAQPAQKQVDIPLRLLVPTQYVGAIIIGKGGATI
RNIKTQTSKIDVHRKENAGAERAISVHSTPEGSSACKMILETHMRKEADTKTAD
VPLKTLAHNFVRLIGREGRNLKKEODTETKITISSLODLTLNPERTITVKAIE
NCCRAEQEIMKKVREAVENDVAAMSLOSHTPGLNLAAGLFPASSAVPPPPSVTG
AAYSSPQWQAEQWQVFIQAQAVGAILIHKGHQHIKLOKHFASASIKIAPPEPDSK
VRMVIITPPPEAQFQAGRIYGLKLEENFFGPKKEVKLETHIRVPASAAGRVIGGGK
TVNELQNLTAEEVVVPRDQTPDENQDVIVIKIIGHFYASQAQRKIRDILAQVKQHQK
GQSQAQAARRK"
```

BASE COUNT 596 a 687 c 664 g 434 t
ORIGIN

alignment_scores:

Quality: 2181.00 Length: 585
Ratio: 4.154 Gaps: 8
Percent Similarity: 89.744 Percent Identity: 74.017

alignment_block:

US-09-685-696-176 x AF198254 ..

Align seg 1/1 to: AF198254 from: 1 to: 2381

```
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
306 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGGCTGACCCCGCGGA 355
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
356 CTTGGAGAAAGTGTTCGGGAGCACAAGATCTCTACAGCGCGCAGTCT 405
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
406 TGGTCAATCCGGCTACGCCTTCGTGGACTCCCGGACGAGCAGCTGGCG 455
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
456 ATGAGGCCATCGAAATCTTCTCCGGGAAAGTAGAATTACAAGGAAACG 505
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
506 CTTAGAGATTGAACATTGCGTGCCTCAAAAAAACAAGGAGCGCGAAATTC 555
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
556 AAATCGAATATTTCCACCCAGCTCCGATGGGAAGTACTGGACAGCTG 605
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
606 CTGGCTCAGTATGTCAGTAGAGAACTGTGAGCAAGTGAACACCGAGAG 655
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
656 TGAGAGCGCAGTGGTGAATGTCACCTATTCCACCGGGGAGCAGACCGG 705
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
706 AACCCATCATGAGCTGAATGATGACCATGTTGGAGAACCATGCTCCCTGA 755
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
756 GTCTCTACATCCCGATGAGCAGATAGCA...CAGGACCTGAGATGG 802
167 nProArgGlyArgArgGlyLeuGlnArgGlySerArgGlnGlys 184
|||||
803 GCGCCGAGGG...GGCTTTGGCTCTCGGGGTTCAGCCCGCCGAGGCT 846
```


SPSPQRAQRGDHSSRQGHAPGCTSQARQIDFPLRLVPTQFVGAIGKEGLTKNI
TKQTSRVDTHRKENSAAEKPTIHAPEGTSEACRMILEIMQKEDETKLAEEIPL
KILAHNGVRLIKEGRENKLIKIEHETGKTITISSLODSLFINPERITVKGTVEACA
SAEIEIMKLRKEAFENMLAVNOQANLIPGLNLSALGTFSTGLSVLPPAGPRGAPPA
APYHPTKHSYFSSLYPHQGFPPPHHSHYPEQEIYNLPIPTQAVGAILCKGKAHLK
QLARFAGASIKIAPAGPDSERVNIITGPPEAKQKGRIFGKLKEENFNPKKEVK
LEAHIRPSSFTGKRGKTKVNELQNLTSAEVIVPRDQTPDENEVIVRIIGHFFA
SQAQRKIREIVQKQOQBOKYPOGVASQSRK*
BASE COUNT 565 a 552 c 532 g 361 t
ORIGIN

alignment_scores:
Quality: 1950.50 Length: 591
Ratio: 3.817 Gaps: 8
Percent Similarity: 86.464 Percent Identity: 65.313

alignment_block:

US-09-685-696-176 x AF117107 ..

Align seg 1/1 to: AF117107 from: 1 to: 2010

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
50 ATGAACAAGCTTTACATCGGGAACCTTGAGCCCGCGCTCACCCGCCGACGA 99
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
100 CCTCCGGCAGCTCTTGGGACAGGAGCTGCCCTGGCGGGACAGCTCC 149
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
150 TCGTGAAGTCGGCTACGCTTCTGGTGGACTACCCCGCAGCAGAACTGGGCC 199
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
200 ATCCGCGCCATCGAGACCTCTCGGGAAGTGAATTCATGGGAAAAAT 249
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
250 CATGGGAAGTTGATTACTAGTCTCTAAAAGCTAAGGACGAGGAAATTC 299
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
300 AGATTTCGAACATCCCTCTCACCCTGAGTGGGAGGTGTGGATGCATCT 349
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
350 TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAC 399
117 rGluThrAlaValAlaValAsnValThrTyrSerSerLysAspGlnAlaArg 134
400 AGNAACCGCGTGTGTCACGTACATATGCACAGAGAAGAACGACAAA 449
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
450 TAGCCATGGAGAAGCTAAGCGGSCATCAGTTTCAGAACTACTCCTTCAAG 499
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
500 ATTTCTATACATCCCGATGAAGAGGTGAGCTCCCTTCGCGCCCTCAG.. 547
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly.. 183
548 ...CGAGCCCGACG...GGGACCACTCTTCCCGGAGCAAGGCC 587
184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
588 ACGCCCTGGGCGACTTCTCAGGCAGACAGATTGATTCCCGCTCGCG 637
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
638 ATCCTGGTCCCGACCCAGTTTGTGTGGCCATCATCGGAAGGAGGCGTT 687

216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
688 GACCAATAAAGAACATCATTAAAGCAGACCCAGTCCCGGGTAGATATCCATA 737
233 rGlyGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
738 GAAAGAGAACTCTGGAGCTGCAGAGAAGCTGTCCACCATCCATGCCACC 787
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMethHisL 266
788 CCAGAGGAGACTTCTGAGCATGCCGATCATCTTGAATCATGAGAA 837
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
838 AGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCTCTCGAAATCTTGG 887
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
888 CACACAATGGCTTGGTGGAGAGCTGATTGGAAAAGAGCAGAAATTTG 937
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG 316
938 AGAAAATTGAACATGAACAGGAGGACCAAGATAACAATCTCATCTTTGCA 987
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
988 GGATTTCAGCATATACACCCGGAAGAACCATCCTGTGAAGGGCACAG 1037
333 alGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgGlu 349
1038 TTGAGGCTGTGCGAGTGTGAGATAGAGATTATGAAGAACCTGCGGTGAG 1087
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1088 GCCTTTGAAAAATGATGTGCTGTAAACCAACCAAGCCCAATCTGATCC 1137
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe..... 376
1138 AGGTTGAACCTCAGCGCATCTGGCATCTTTCAACAGGACTGTCCGTGC 1187
377ProThrSer.....GlyMetProProThr..... 386
1188 TATCTCCACGAGCGGCCCGGAGCTCCCGCGCTGCCCTTACCAC 1237
387SerGlyProProSerAlaMetThrPro..... 395
1238 CCCTTCACCTACCCACTCCGATATCTTCCAGCTGTACCCCTCACCA 1287
396ProTyrProGlnPheGluGln...SerGluThrGluThrValH 409
1288 GTTGGCCCGTTCCTCCGATCATCTCTTATCCAGAGCAGGAGATTGTGA 1337
409 isGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425
1338 ATCTCTTATCCCAACCCAGGCTGTGGGCCCATCATCGGGAAGAAGGG 1387
426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAl 442
1388 GCACACATCAACAGCTGGCGAGATTGCGCGGAGCCTCTATCAAGATTGC 1437
442 aProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyP 459
1438 CCTCGGAGGCGCCAGCTCAGCGAAAGGATGGTCATCATCACCAGGC 1487
459 roProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLys 475
1488 CACCGGAAGCCAGTTCAAGGCCCGGAGCGGATCTTTGGGAAACATGAA 1537
476 GluGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisI 492
1538 GAGGAAACTTCTTTAACCCCAAGAGAAAGTGAAGCTGAAGCGCATAT 1587

492 eArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysT 509
|||||
1588 CAGAGTGGCCCTCTCCACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGA 1637
|||||
509 hrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArg 525
|||||
1638 CCGTGAACGAACCTGCAGAACTTAACCACTGCAGAGTCATCGTGCCTCGT 1687
|||||
526 AspGlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHi 542
|||||
1688 GACCAACGCCAGATGAAATGAGGAAGTGTGTCAGAAATTATCGGGCA 1737
|||||
542 sPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrG 559
|||||
1738 CTTCTTTGCTAGCCCACTGCACAGCCGCAAGATCAGGGAAATTGTACAAC 1787
|||||
559 lnValLysGlnHisGlnGln 566
|||||
1788 AGGTGAACGACGAGGAGAGAAA 1810
|||||

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DE KH DOMAIN-CONTAINING TRANSCRIPTION FACTOR B3.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92249652; PubMed=1577195;
 RA Pfaff S.L., Taylor W.L.;
 RT "Characterization of a Xenopus oocyte factor that binds to a
 RT developmentally regulated cis-element in the TFIIB gene."
 RL Dev. Biol. 151:306-316(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Griffin D., Taylor W.L.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
 RA Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
 RL Genes Dev. 0:0-0(1998).
 DR EMBL; AF042353; AAB97457.1; -.
 DR EMBL; AF064633; AAC18597.1; -.
 DR InterPro; IPR000504; RRM.
 DR InterPro; IPR000958; KH.
 DR Pfam; PF00013; KH-domain; 4.
 DR Pfam; PF00076; rrm; 2.
 DR PROSITE; PS50102; RRM; 2.
 DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 SQ SEQUENCE 593 AA; 65385 MW; 5A5AB4B4A1D55DF7 CRC64;

 Query Match 84.0%; Score 2483; DB 13; Length 593;
 Best Local Similarity 82.1%; Pred. No. 1.5e-160;
 Matches 487; Conservative 47; Mismatches 45; Indels 14; Gaps 6;

 QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKAEALSCK 60
 DB 1 MNKLYIGNLSENVSPDLESIFKESKIPFTGQFLVKSQYAFVDCPDSEWALKAEALSCK 60

 QY 61 IELHCKPIEVESHVPKRQIRKLOIRNIPPHLOWEVLDSLLVQGVVSECEQVNTDSETA 120
 DB 61 VELHCKVIEVESHVPKRQIRKLOIRNIPPHLOWEVLDSLLVQGVVSECEQVNTDSETA 120

 QY 121 VVNVYSSKQDQARQALDKLNGFQLENFTLKVAIIPDEMAAQNPLQQ-----PRGR 172
 DB 121 VVNVYANKEHARQGLEKNGYQLENYSKVTYIPDEMATPSPSQLOOQPOQHPOGR 180

 QY 173 GLGGRSSRQSGPSVSKOKP-CDLPLRLVPTQFVGAIGKEGATIRNITKQTSKIDV 231
 DB 181 GFGGRPARQSGPAAARPKQSEVPLRMLVPTQFVGAIGKEGATIRNITKQTSKIDI 240

 QY 232 HRKENAGAAEKSTILSTPEGTSACKSILEIMHKEAODKFTTEIPLKILAHNFVGR 291
 DB 241 HRKENAGAAEKPTITHSTPEGSAACKIIMEIMQEAODKFTTEIPLKILAHNFVGR 300

 QY 292 IGKEGRNLKTEQDDTKITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRSY 351
 DB 301 IGKEGRNLKTEQDDTKITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRSY 360

 QY 352 ENDIASMNLQALHLPGLNLNGLFPP-TSGMPPTTSGPPPSAMT-PPYPQF-EQSETETV 408
 DB 361 ENDIAAMNLQALHLPGLNLNGLFPPSSSGMPPPSAGVSSPTTSASYPFQSQPESETV 420

 QY 409 HOFIPALSGAIGQGOHIKQLSRFAGASIKIAPAEADPAKVRVITGPPPEAFQKAG 468
 DB 421 HLFIPALVGAIGQGOHIKQLSRFAGASIKIAPAEADPAKVRVITGPPPEAFQKAG 480

QY 469 RIYGIKEENFVSPKEVKEAHIRVPSFAAGRVIGKGTVNELONLSSAEVVVPRDQT 528
 DB 481 RIYGIKEENFVSPKEVKEAHIRVPSFAAGRVIGKGTVNELONLSSAEVVVPRDQT 540

 QY 529 PDNDVVVVIKTHFVACQVARKTQELITQVK--OHQOQKALOSGPPQSRRK 579
 DB 541 PDNDVVVVIKTHFVACQVARKTQELITQVK--OHQOQKALOSGPPQSRRK 593

 RESULT 4
 O73932 PRELIMINARY; PRT; 594 AA.
 AC O73932;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE VGI RNA BINDING PROTEIN VARIANT D.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
 RA Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
 RL Genes Dev. 0:0-0(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98228351; PubMed=9560341;
 RA Deshler J.O., Highett M.I., Abramson T., Schnapp B.J.;
 RT "A highly conserved RNA-binding protein for cytoplasmic mRNA
 RT localization in vertebrates."
 RL Curr. Biol. 8:489-496(1998).
 DR EMBL; AF064634; AAC18598.1; -.
 DR EMBL; AF055923; AAC41285.1; -.
 DR InterPro; IPR000504; RRM.
 DR InterPro; IPR000958; KH.
 DR Pfam; PF00013; KH-domain; 4.
 DR Pfam; PF00076; rrm; 2.
 DR PROSITE; PS50102; RRM; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 SQ SEQUENCE 594 AA; 65643 MW; 54CEA7BFF0856DD6 CRC64;

 Query Match 83.3%; Score 2461.5; DB 13; Length 594;
 Best Local Similarity 81.3%; Pred. No. 4.4e-159;
 Matches 483; Conservative 49; Mismatches 47; Indels 15; Gaps 6;

 QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKAEALSCK 60
 DB 1 MNKLYIGNLSENVSPDLESIFKESKIPFTGQFLVKSQYAFVDCPDSEWALKAEALSCK 60

 QY 61 IELHCKPIEVESHVPKRQIRKLOIRNIPPHLOWEVLDSLLVQGVVSECEQVNTDSETA 120
 DB 61 VELHCKVIEVESHVPKRQIRKLOIRNIPPHLOWEVLDSLLVQGVVSECEQVNTDSETA 120

 QY 121 VVNVYSSKQDQARQALDKLNGFQLENFTLKVAIIPDEMAAQNPLQQ-----PRGR 171
 DB 121 VVNVYANKEHARQGLEKNGYQLENYSKVTYIPDEMATPSPSQLOOQPOQHPOGR 180

 QY 172 RGLGGRSSRQSGPSVSKOKP-CDLPLRLVPTQFVGAIGKEGATIRNITKQTSKID 230
 DB 181 GFGGRPARQSGPAAARPKQSEVPLRMLVPTQFVGAIGKEGATIRNITKQTSKIDI 240

 QY 231 HRKENAGAAEKSTILSTPEGTSACKSILEIMHKEAODKFTTEIPLKILAHNFVGR 290
 DB 241 HRKENAGAAEKPTITHSTPEGSAACKIIMEIMQEAODKFTTEIPLKILAHNFVGR 300

 QY 291 LGKEGRNLKTEQDDTKITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRSY 350
 DB 301 LGKEGRNLKTEQDDTKITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRSY 360

```
QY 351 YENDIASMNLQAHILPGLNUNALGLFP-PTSGMPPPTSGPPS-AMTPPYQF-EQSTET 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 YENDIAAMNLQAHILPGLNUNALGLFPSSSGMPPSGVPSPTSSYPPFGQOPESET 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 408 VHOFPALSGAIIKGOGHIKOLSRFAGASIKIAPAEAPDAKVRWIIITGPPPEAQK 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 VHLFPALAVGAIIGOGHIKOLSRFAGASIKIAPAEAGPDALRMVIIITGPPPEAQK 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 468 GRIYGIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVPDQ 527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 GRIYGLKEENFPGKEEVKLEHIRVPSYAAGRVIGKGGKTVNELQNLSSAEVVPDQ 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 528 TPDENDVVYKITGHFYACQVQARKIQEILITQVK--QHQQOKALQSGPPPSRRK 579
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 TPDENDVVYKITGHFYASQLAQKIQEILAQVRRQOQOKTVQSGOPQPRRK 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
Q9PW80 PRELIMINARY; PRT; 582 AA.
ID Q9PW80;
AC Q9PW80;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Vcl RNA BINDING PROTEIN.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA Taylor W., Meyer D., Standart N., Raz E., Visraeli J.K.;
RA "vcl RBP intracellular distribution and evolutionarily conserved
RT expression suggest multiple roles during development.";
RL Mech. Dev. 0:0-(1999).
DR EMBL: AF161270; AAD45610.1;
DR InterPro: IPR000504; RRM.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS50102; RRM; 2.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
SQ SEQUENCE 582 AA; 63351 MW; 9DA563200681B306 CRC64;
```

```
Query Match 76.5%; Score 2261.5; DB 13; Length 582;
Best Local Similarity 77.1%; Pred. No. 1.6e-145;
Matches 450; Conservative 50; Mismatches 77; Indels 7; Gaps 5;
```

```
QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKATEALSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MNKLYIGNVSEASALDESIFEQWKIPSPAPFLVSGYAFVDCPDSEWALKATEALSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 IELHGKPIEVESHVPKQRIRKLIQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VELHGVLEVEHVPKQRIRKLIQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 VNVVTSKDKQARQALDKNGFQLENFTLKVAYIPDEMAAQNPLOQPRRRGLGQRSS 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNVVRYCAKDQAREAMDKNGLFMENVALKVSYPIDETAADAP--AVGRRGPNRGP 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 ROGSPGSVKQK-PCDLPRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKENA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 RQSPSLGAPRKLDQVPLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKENA 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 240 AKSITILSTPEGTSACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRLLGKEGRNL 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 AKRPITVHSTPEGSSACRNIMEIQKEADTKITEIPLKILAHNNFVGRLLGKEGRNL 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 300 KKIEQDTRDKITITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRESYENDIASMN 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 KKIEQDTRDKITITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRESYENDIAAMH 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 360 LOAHLIPGLNUNALGLFP-PTSG--MPPTSGPPSAMTPPYQF-EQSETETVHQFIPAL 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 LQSNLIPGLNUNALGLFPGAASGGSISVSVGGPPGAAGYQSGFQAQWSETHLFLIPAL 418
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 416 SVGAIIKGOGHIKOLSRFAGASIKIAPAEAPDAKVRWIIITGPPPEAQKAGRIYGIK 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 AVGAIIGOGHIKOLSRFAGASIKIAPADIDAKRMIISGPPPEAQKAGRIYGIK 478
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 476 EENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVPDQTPDENDQV 535
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 EENFFGPKKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVPDQTPDENDQV 538
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 536 VVKITGHFYACQVQARKIQEILITQVKHQOQOKALQSGPPPSRRK 579
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 VVKITGHFYASQLAQKIQEILISQVRRQOQPKPSAAGPPVARRK 582
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 6
Q42254 PRELIMINARY; PRT; 576 AA.
ID Q42254;
AC Q42254;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ZIPCODE-BINDING PROTEIN.
GN ZAP1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97220007; PubMed=9121465;
RA Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL Mol. Cell. Biol. 17:2158-2165(1997).
DR EMBL: AF026527; AAB82295.1;
DR InterPro: IPR000504; RRM.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS50102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
SQ SEQUENCE 576 AA; 63271 MW; 01AAF2D1D81C8811 CRC64;
```

```
Query Match 75.1%; Score 2219.5; DB 13; Length 576;
Best Local Similarity 74.7%; Pred. No. 1.2e-142;
Matches 437; Conservative 61; Mismatches 72; Indels 15; Gaps 7;
```

```
QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKATEALSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MNKLYIGNLSVETPADLEKVFNDHKISFSQFLVSGYAFVDCPDSEWALKATEALSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 IELHGKPIEVESHVPKQRIRKLIQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VELHGVLEVEHVPKQRIRKLIQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 VNVVTSKDKQARQALDKNGFQLENFTLKVAYIPDEMAAQNPLOQPRRRGLGQRSS 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNVVTSKDKQARQALDKNGFQLENFTLKVAYIPDEMAAQNPLOQPRRRGLGQRSS 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 180 ROGSP---GSVSKQPCDLPRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKEN 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 177 PROGSPVAPVKKQPVDPILRLVPTQVVGAIIGKEGATIRNITKQTSKIDVHRKEN 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Qy 237 AGAAEKSTILSTPGCTSAACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRIGKEG 296
 Db 237 AGAAEKAIISHTPGCSAAKMLEIMHKEAKDTKTADVEPLKILAHNNFVGRIGKEG 296
 Qy 297 RNLLKIEODTQTKTISPQLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Db 297 RNLLKIEODTQTKTISPQLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Qy 357 SMNLQAHILPGLNLAALGFLPPTSGMPPPTSGPPSAMT--PPYQFQESQETVHQFIPA 414
 Db 357 AMSLQSHLIPGLNLAALGFLPASPNAVPP---PPSSVGAAPYSFMPPEQETVHVFI 413
 Qy 415 LSVGAIGKQGHQIKQLSRFAGASIKIAPAPDAKVRWVIITGPPEAQFRAQRIYKGI 474
 Db 414 QAVGAIGKQGHQIKQLSRFASIKIAPPETPDSKVRWVITGPPEAQFRAQRIYKGL 473
 Qy 475 KEENFVSPKEVKLEAHLRVPFSAAGRVIGKGTVNELQNLSSAEVVVPDQTPDENQ 534
 Db 474 KEENFVSPKEVKLEAHLRVPFSAAGRVIGKGTVNELQNLSSAEVVVPDQTPDENQ 533
 Qy 535 VVVKITGHFYACQVAQRKIQIILTVQKQHQKQKALQSGPPQSRK 579
 Db 534 VIVKIIGHFYASQMAQRKIRILAQVKQ-QHKG-QSGQLQARRK 576

RESULT 7
 O88477 PRELIMINARY; PRT; 577 AA.
 AC O88477;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CODING REGION DETERMINANT BINDING PROTEIN (CODING REGION DETERMINANT-BINDING PROTEIN).
 GN CRDBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92217743; PubMed=1559612;
 RA Bernick P.L., Herrick D.J., Prokipcak R.D., Ross J.;
 RT "Control of c-myc mRNA half-life in vitro by a protein capable of
 RL binding to a coding region stability determinant.";
 RL Genes Dev. 6:642-654(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9415886; PubMed=8114742;
 RA Herrick D.J., Ross J.;
 RT "The half-life of c-myc mRNA in growing and serum-stimulated cells:
 RL influence of the coding and 3' untranslated regions and role of
 RT ribosome translocation.";
 RL Mol. Cell. Biol. 14:2119-2128(1994).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94179348; PubMed=8132663;
 RA Prokipcak R.D., Herrick D.J., Ross J.;
 RT "Purification and properties of a protein that binds to the C-terminal
 RL coding region of human c-myc mRNA.";
 RL J. Biol. Chem. 269:9261-9269(1994).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97322234; PubMed=9178888;
 RA Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
 RA Gruppiso P.A., Ross J.;
 RT "Developmental regulation of CRD-BP, an RNA-binding protein that
 RL stabilizes c-myc mRNA in vitro.";
 RL Oncogene 14:1279-1286(1997).
 [5]
 RP SEQUENCE FROM N.A.
 RX Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
 RA Fleisig A.J.;

Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AF061569; AAC72743.1;
 DR EMBL; AK013940; BAB29071.1;
 DR MGD; MGI:1330862; Crdbp.
 DR InterPro; IPR000958; KH.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00013; KH-domain; 4.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS0102; RRM; 2.
 SQ SEQUENCE 577 AA; 63450 MW; EFB1A2FF9F0344 CRC64;

Query Match 74.2%; Score 2192; DB 11; Length 577;
 Best Local Similarity 74.1%; Pred. No. 8.5e-141;
 Matches 434; Conservative 62; Mismatches 74; Indels 16; Gaps 8;

Qy 1 MNKLYIGNLSENAFSDLESIFPKDAKIPVSGPLVKTGYAFVDCPDSEWALKAEIALSK 60
 Db 1 MNKLYIGNLSEVTPADLEKVFAEHKISYSGQFLVKGSAFVDCPDSEWALKAEIALSK 60
 Qy 61 IELHCKPIEVEHSPKRIKQIRNIPPHLOVEVLDLVLVQGVESCEQVNTDSETA 120
 Db 61 VELOKRLIEHSPKRSRRIQIRNIPPHLOVEVLDLVLVQGVESCEQVNTDSETA 120
 Qy 121 VVNVTVSSKQDARALDKINGFLENFTLVKVAIPDEMAAQNPQQPRRR-GLGQREGS 179
 Db 121 VVNVTVSSNREQTRQAIMKLNHQLNHALKVSIIPEQITQ---GPNRGRGGFGRSQ 176
 Qy 180 SRQSGP---GSVSKQKPCDPLRLVLVPTQFVGAIIKEGATIRNITKQTSKIDVHRKEN 236
 Db 177 PRQSPVAAGAPAKQOPVDIPLRLVLVPTQVGAIIKEGATIRNITKQTSKIDVHRKEN 236
 Qy 237 AGAAEKSTILSTPGCTSAACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRIGKEG 296
 Db 237 AGAAEKAIISHTPGCSAAKMLEIMHKEAKDTKTADVEPLKILAHNNFVGRIGKEG 296
 Qy 297 RNLLKIEODTQTKTISPQLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Db 297 RNLLKIEODTQTKTISPQLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Qy 357 SMNLQAHILPGLNLAALGFLPPTSGMPPPTSGPPSAMT--PPYQFQESQETVHQFIPA 413
 Db 357 AMSLQSHLIPGLNLAALGFLPASPNAVPP---PPSSVGAAPYSFMPPEQETVHVFI 413
 Qy 414 ALSVGAIGKQGHQIKQLSRFAGASIKIAPAPDAKVRWVIITGPPEAQFRAQRIYK 473
 Db 414 QAVGAIGKQGHQIKQLSRFASIKIAPPETPDSKVRWVITGPPEAQFRAQRIYK 473

```
QY 474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 474 LKEENFPGPKEEVKLETHIRVPASAGRVIGKGGKTVNELQNLTAEEVVVPRDQTPDEND 533

QY 534 QVVVKITGHFYACOVAQRKIQTIELITQVKKHQKQKALQSGPPQSRKK 579
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 534 QVIVKIIGHFYASOMAQRKIRDLAQVKQ-OHQKG-QSNLAQARRK 577

RESULT 8
Q9NZI8 PRELIMINARY; PRT; 577 AA.
ID Q9NZI8
AC Q9NZI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MRNA-BINDING PROTEIN CRDBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA Panoutsakopoulos G., Kyriazoglou I., Voutzoulis S., Tsiapalis C.M.,
RA Kittas C., Agnantis N., Pandis N.;
RT "Ectopic expression of a KH-domain containing protein, highly
RT homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT malignant mesenchymal tumors.";
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF198254; AAF37203.1; -.
DR InterPro: IPR000958; KH.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 577 AA; 63456 MW; 0749A060F252D81D CRC64;
```

```
Query Match 73.8%; Score 2181; DB 4; Length 577;
Best Local Similarity 73.9%; Pred. No. 4.8e-140;
Matches 433; Conservative 61; Mismatches 76; Indels 16; Gaps 8;
```

```
QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKAEALSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MNKLYIGNLSENVTPADLEKVFAEHKISYSGQFLVKSQGYAFVDCPDSEWALKAEALSGK 60

QY 61 IELHGKPIEVESHVSPKQRIRKQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VLOGKRLEIHSVPPKQSRKQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120

QY 121 VNVVTSYSSKDOARQALDKLNGFQLENFTLKVAYIPDEMAQQNPLOQPRRRGLQGRGS 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNVVTSYSSKDOARQALDKLNGFQLENFTLKVAYIPDEMAQQNPLOQPRRRGLQGRGS 176

QY 180 SROGSP---GSVSKQPCDLPRLRLVPTQFVGAIIGKEGATIRNITQSKIDVHRKEN 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 177 PROGSVAAGAPAKQOQVDPLELLVPTQFVGAIIGKEGATIRNITQSKIDVHRKEN 236

QY 237 AGAAEKSIITLSPETSACKSILEIMHKEADIKFTEIPLKILAHNNFVGRLLGKEG 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 237 AGAAEKASIVHSITPEGSSACKMILEIMHKEADTKTAEVPLKTLAHNNFVGRLLGKEG 296

QY 297 RNLKKIEQDFTKTIISPLQELTLYNPRTITVKGNVETCAKAEIEIMKKIRESYENDIA 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 297 RNLKKVEQDFTKTIISPLQELTLYNPRTITVKGNAIENCCRAEQEIMKKVREAYENDVA 356

QY 357 SNLQAHILPGLNLALGLFPPTSGMPPTSGPPSAMT---PPYPQEQS-ETETVHQFIP 413
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 357 ANSLQHLTPGLNLAAVGLFPASSAVPP---PPSVTGAAPYSPPMQAPEQEMVQVFP 413
```

```
QY 414 ALSVGAIIGKQGHRIKOLSRFAGASIKIAPAEAPDAKVRMVIITGPPPEAQKAGRIYKG 473
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 AQAVGAIIGKGGHRIKOLSRFASASIKIAPETPDSKVRMVIITGPPPEAQKAGRIYKG 473

QY 474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 474 LKEENFPGPKEEVKLETHIRVPASAGRVIGKGGKTVNELQNLTAEEVVVPRDQTPDEND 533

QY 534 QVVVKITGHFYACOVAQRKIQTIELITQVKKHQKQKALQSGPPQSRKK 579
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 534 QVIVKIIGHFYASOMAQRKIRDLAQVKQ-OHQKG-QSNLAQARRK 577

RESULT 9
Q9Y6M1 PRELIMINARY; PRT; 556 AA.
ID Q9Y6M1
AC Q9Y6M1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HEPATOCELLULAR CARCINOMA AUTOANTIGEN.
GN P62.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99207072; PubMed-10190901;
RA Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
RT in human hepatocellular carcinoma.";
RL J. Exp. Med. 189:1101-1110(1999).
DR EMBL; AF057352; AAD31596.1; -.
DR InterPro; IPR000504; RRM.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR PROSITE; PS50102; RRM; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
SQ SEQUENCE 556 AA; 61842 MW; 1EDEFB100443DDC4 CRC64;
```

```
Query Match 63.2%; Score 1868; DB 4; Length 556;
Best Local Similarity 64.8%; Pred. No. 8.1e-119;
Matches 368; Conservative 78; Mismatches 96; Indels 26; Gaps 6;
```

```
QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKAEALSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 MNKLYIGNLSPAVTADLRQLFGDKRLPLAGQVLLKSGYAFVDPQONWAIETLSGK 61

QY 61 IELHGKPIEVESHVSPKQRIRKQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 VELHGMIEVDYSVSKLSRKQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 121

QY 121 VNVVTSYSSKDOARQALDKLNGFQLENFTLKVAYIPDEMAQQNPLOQPRRRGLQGRGS 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 VNVVTSYSSKDOARQALDKLNGFQLENFTLKVAYIPDEMAQQNPLOQPRRRGLQGRGS 177

QY 181 ROG-SPGVSQKQPCDLPRLRLVPTQFVGAIIGKEGATIRNITQSKIDVHRKENAGA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 EQHAFGGTTSQARQIDFPLRLVPTQFVGAIIGKEGATIRNITQSKIDVHRKENAGA 237

QY 240 AKSITPILTPETSACKSILEIMHKEADIKFTEIPLKILAHNNFVGRLLGKEGRNL 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 AKPVTIHTATPEGTSEACRMILEIMHKEADETKLAIEPLKILAHNNFVGRLLGKEGRNL 297

QY 300 KKEQDFTKTIISPLQELTLYNPRTITVKGNVETCAKAEIEIMKKIRESYENDIASMN 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 298 KKEHETGKTIISPLQELTLYNPRTITVKGNVETCAKAEIEIMKKIRESYENDIASMN 357
```


[illegible]

| | |
|--------|--|
| RESULT | 15 |
| Q23487 | |
| ID | PRELIMINARY; PRT; 768 AA. |
| AC | Q23487; |
| DT | 01-NOV-1996 (TrEMBLrel. 01, Created) |
| DT | 01-NOV-1996 (TrEMBLrel. 01, Last sequence update) |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Last annotation update) |
| DE | SIMILAR TO YEAST PROTEIN HX AND E. COLI PNP. |
| OS | ZK418.9. |
| GN | Caenorhabditis elegans. |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea; |
| OC | Rhabditiidae; Peloderinae; Caenorhabditis. |
| OX | NCBI_TaxID=6239; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=BRISTOL N2; |
| RX | MEDLINE=94150718; PubMed=7906398; |
| RA | Wilson R., Ainscough R., Andersson K., Baynes C., Berks M., |
| RA | Bonfield J., Burton J., Connell M., Copsey T., Coulson A., |
| RA | Crafton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., |
| RA | Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., |
| RA | Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., |
| RA | Liftoning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., |
| RA | Pearsons J., Percov C., Rifkin L., Roopra A., Saunders D., Showkneen R., |

RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: U00047: AAA50693.1: -;
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR000536; Hormone_rec_lig.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00104; hormone_rec; 1.
DR PROSITE: PS00402; BPD_TRANS_INN_MEMBER; UNKNOWN_1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00322; KH; 4.
SQ SEQUENCE 768 AA; 83886 MW; 3672FC807710F20A CRC64;

Query Match 8.2%; Score 243.5; DB 5; Length 768;
Best Local Similarity 21.2%; Pred. No. 2.6e-08;
Matches 123; Conservative 103; Mismatches 215; Indels 139; Gaps 22;
QY 62 ELHGKPIEVEHSPKQRIKRLQIRNIPPHLQWEVLDSLVLQYGVVESCQVNTDSETAV 121
DB 139 QLIGKLIELD--ITKNESL-----LLSALFL-----CSAVHPDISTEG 174
QY 122 VNTYSSKDAQAL-----DKLNGFQ-LENFTLKVAYIPDEMAAQON 163
DB 175 KNLLYKYYQVYSAALINHCCLTNQONAPRSYSELLSLYQIIETHQKIIIGADGGIGTLKR 234
QY 164 PLOO-----PRGRGLQGRSSRGSPGSVSKQKPCDPLPLLVPTQFVGAIGKEG 215
DB 235 PLQSEILDGLIPTKKS--SEVGLNMGDSKITDIYP-----VPEKVVGLVIGKG 284
QY 216 ATIRNITKQTSKIDVHRKENAGAAEKSIITLSTPETSAAKKSILEIMHKE---AODIK 272
DB 285 SEIRLIQOTSGCRVOMDPDQHSVNGFNCTIEGPPQVAVAROMITOVINRNQTGAQPGA 344
QY 273 FTEIPIKILAHNNFVGRLLIGKEGRNLKKEQTDITKITISPLQELTLY----NPERTIT 328
DB 345 APGEVTEMLIPADKIGLVIGKGETIRIVQEQSGLR-NCNVVQETTATGQPKPLRMIG 403
QY 329 VGNVETCAKAEIEIMKKIRESYENDIASMNLQAHLIPLGLNLNALGLFPPTSGMPPTSG 388
DB 404 SPAAIET-ATA-----LVHNMNNTQGNAPLLQRAHQPSG 438
QY 389 PPSAMTPPYQF-----EQSETETVHQFIPALSVGAIIGKOGOHIKOLSRFAGASIK 440
DB 439 -----QFGGGYCAQEAQAKGEVI---VPLRSLNGMIIGKGGEMIKKLAATGTGKIQ 485
QY 441 IAPAEAPDAKRVVITGPPEAFKAQRIYGIKE--ENFVSPKEVKLEA----- 490
DB 486 FKPDTPNSEDRIAVINGTRDQIYRATERITEIVNRKKNAGAPQDRGSAGTVLPQOSIF 545
QY 491 HIRVPFAAGRVIGKCKTVNELQNLSSAEV-VVPRDQTPDENDQVVVKITGH-----FYA 545
DB 546 YMHVPAGKCGLVIGKGENIKITERTGTACGLAPAAEQKNE-DEKVFEEKSGLQIHTA 604
QY 546 CQVAQRKIQEILTQVK-----HQOQKALQSGPPQS 576
DB 605 SHLVRIKVGELSPNTVPVPLQAGAGGYQOQOAMFSAGTQN 644

Search completed: January 25, 2002, 22:19:19
Job time: 440 sec

OM of: US-09-685-696-176 to: Issued_Patents_NA:* out_format : pfs
 Date: Jan 25, 2002 11:22 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=frame+p2n_model -DEV=xlp
 -O=/cgn2_1/USPTO.spool/US09685696/runat_25012002.145551.10660/app_query.fasta_1.648
 -DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09685696 -CGN1_1_87 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
 -WAIT -THREADS=1

Search information block:
 Query: US-09-685-696-176
 Query length: 579
 Database: Issued_Patents_NA:*
 Database sequences: 351203
 Database length: 11323899
 Search time (sec): 90.080000

score_list:

| Sequence | Strd Orig | zscore | Escore | Len | Documentation |
|--|-----------|---------|---------|------|---------------|
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-061-709-4 | .. | 2956.00 | 5412.74 | 4159 | 7.7e-294 |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-261-855-1 | .. | 2192.00 | 4009.61 | 2224 | 1.1e-215 |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-6 | .. | 1950.50 | 3538.22 | 3412 | 1.5e-190 |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-7 | .. | 1873.00 | 3422.14 | 1946 | 5.8e-183 |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-8 | .. | 1868.00 | 3406.34 | 4283 | 4.4e-182 |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-5 | .. | 1637.00 | 2987.92 | 1708 | 8.9e-159 |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-021-608D-1 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-1 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-2 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-3 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-4 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-5 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-6 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-7 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-8 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-9 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-10 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-11 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-12 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-13 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-14 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-15 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-16 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-17 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-18 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-19 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-20 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-21 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-22 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-23 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-24 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-25 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-26 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-27 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-28 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-29 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-30 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-31 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-32 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-33 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-34 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-35 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-36 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-37 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-38 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-39 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-40 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-41 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-42 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-43 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-44 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-45 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-46 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-47 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-48 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-49 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-50 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-51 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-52 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-53 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-54 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-55 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-56 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-57 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-58 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-59 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-60 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-61 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-62 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-63 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-64 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-65 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-66 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-67 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-68 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-69 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-70 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-71 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-72 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-73 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-74 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-75 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-76 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-77 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-78 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-79 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-80 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-81 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-82 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-83 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-84 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-85 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-86 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-87 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-88 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-89 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-90 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-91 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-92 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-93 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-94 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-95 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-96 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-97 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-98 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-99 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-100 | .. | 241.00 | 405.53 | 2384 | 6.1e-15 |

/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-974-4 + 121.50 185.40 0.0112 2277
 /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-974-5 + 121.50 185.40 0.0112 2277
 /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-098-487-4 + 121.50 185.40 0.0112 2277
 /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-098-487-5 + 121.50 185.40 0.0112 2277
 /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-967-3 + 121.50 183.11 0.0150 2733
 seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-4
 seq_documentation_block:
 ; Sequence 4, Application US/09061709B
 ; Patent No. 6297364
 ; GENERAL INFORMATION:
 ; APPLICANT: Chen, Yao-Tseng
 ; APPLICANT: Gure, Ali
 ; APPLICANT: Tsang, Solam
 ; APPLICANT: Stockert, Elisabeth
 ; APPLICANT: Jager, Elke
 ; APPLICANT: Knuth, Alexander
 ; APPLICANT: Old, Lloyd J.
 ; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
 ; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
 ; FILE REFERENCE: LUD 5538
 ; CURRENT APPLICATION NUMBER: US/09/061,709B
 ; CURRENT FILING DATE: 1998-04-17
 ; NUMBER OF SEQ ID NOS: 8
 ; SEQ ID NO 4
 ; LENGTH: 4159
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; US-09-061-709-4

alignment_scores:
 Quality: 2956.00 Length: 579
 Ratio: 5.105 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-685-696-176 x US-09-061-709-4
 Align seg 1/1 to: US-09-061-709-4 from: 1 to: 4159
 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
 |||||
 251 ATGAACAACTGTATATCGAAACCTCAGCAGACGCGCCCTCGGA 300
 17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyPropheL 34
 |||||
 301 CCTAGAAAGTATCTCAAGAGCGCCCAAGATCCCGTCTCGGACCTTCC 350
 34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
 |||||
 351 TGGTGAACACTGGCTACCGCTTCGGAGTCCCGGACGAGAGCTGGGCC 400
 51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
 |||||

601 GGAAACTGCAGTTGTAATGTAACCTATTTCAGTAAGGACCAAGCTAGAC 650
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
151 AAGCACTAGACAACTGAATGATTTTCAGTTAGAGAAATTTACCTTGAAA 700
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
701 GTAGCCTATATCCCTGATGAATATGCGCCCGCCAGCAAAACCCCTTCAGCA 750
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlys 184
751 GCCCGAGTCCCGGGGCTTGGCGACAGGGGCTCCTCAAGCAGGGGT 800
184 exProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
801 CTCCAGGATCCGTATCCCAAGCAGAAACCATGTGATTTGCCCTGCGCCTG 850
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAla 217
851 CTGGTTCCACCAATTTGTTGGAGCCATCATAGAAAAGAGGTGCCAC 900
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
901 CATTCGGAAACATCACCAACAGACCCAGCTTAATAATCGATGTCACCGTA 950
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
951 AAGAAATGCGGGGCTGCTGAGAAAGTCGATTACTATCTCTACTCCT 1000
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG 267
1001 GAAGCACCTCTCGCGCTGTGAAGTCTATTCTGGAGATTATGCATAAGGA 1050
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
1051 AGCTCAAGATATAAATTCACAGAGAGATCCCTTGAAGATTTAGCTC 1100
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
1101 ATAATACTTTGTTGGAGCTCTATTGTTAAAGAAAGGAAATCTTAAA 1150
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
1151 AAAATTGACGACGACAGACACATAAATCACGATATCTCATTCAGGAG 1200
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
1201 ATTGACGCTGTATATCCAGAACGCACTATTACAGTTAAAGGCAATGTTG 1250
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
1251 AGACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATCAGGAGTCT 1300
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG 367
1301 TATGAAATGATATGCTTCTATGAATCTTCAAGCACATTTAATTCCTGG 1350
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
1351 ATTAATCTGAACGCTTGGGCTGTCTCCACCCACTTCAGGGATGCCAC 1400
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
1401 CTCCACCTCAGGGCCCTTTCAGCATGACTCTCCCTACCCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
1451 GAGCAATCAGAAACGGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lclValaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgp 434
1501 CGGTGCCATCATCGCAAGCAGGCGCCAGCACATCAAGCAGCTTCTCGCT 1550

434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
1551 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGGAAGCACCAGATGCTAAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG 467
1601 GTGAGGATGGTATTATCAGTACCAGCAGAGGCTCAGTTCAAGGCTCA 1650
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
1651 CGGAAGAATTTATGCAAAATTAAGAGAAACCTTTGTAGTCTCTAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaLagly 500
1701 AAGAGTGAACCTTGAAGCTCATATCAGAGTGCCATCTCTTCTGCTGCTGC 1750
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1751 AGAGTTATTGGAAGAGGAGGCAAAACGGTGAATGAACCTCAGAATTTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
1801 AAGTGCAGAAAGTTGTTGCTCCCTGTCGACGACACCTGATGAGAAATGACC 1850
534 InValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
1851 AAGTGGTTGCAAAATAACTGGTCACTTCTATGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLys 567
1901 AGAAATTTCAAGAAATTTCTGACTCAGGTAAAGCAGCACCACACAGAA 1950
567 sAlaLeuGlnSerGlyProGlnSerArgArgLys 579
1951 GGCTCTGCAAGTGGACCCCTCAGTCAAGACGGAAG 1987
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-261-855-1
seq_documentation_block:
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/261,855A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2224
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-261-855-1

alignment_scores:
Quality: 2192.00 Length: 585
Ratio: 4.152 Gaps: 8
Percent Similarity: 90.256 Percent Identity: 74.188
alignment_block:
US-09-685-696-176 x US-09-261-855-1 ..
Align seg 1/1 to: US-09-261-855-1 from: 1 to: 2224
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
131 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGTGTGACCCCGCAGA 180
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34

```

181 CTTGGAGAAAGTATTCCGAGGACACAAAGATCTCTACAGCGCCAGTTCT 230
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
231 TGGTCAATCCGGCTACGGCTTCGTGGATTGCCCGCAGCAGCACTGGCG 280
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
281 ATGAAGCCATCGAAATTTCTCGGGAAAGTAGAACTGCAAGGAAACG 330
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
331 TCTAGAGATTGAACACTCAAGTCCCAACAAACAAAGAGAGTCCGAAATAC 380
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
381 AGATCCGCAATATTCCACCTCAGCTCCGATGGGAAGTGTAGATAGCC 430
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
431 CTGGCTCAGTACGGTACAGTGGAGAAGTGTAGCAAGTGAACACTGAAAG 480
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
481 TGAGACAGCGGTGGTCAAGCTCACCCTACTTAACCGGGAGCAGCACGCG 530
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
531 AGCTATCATGAAGCTAATAGGCCATCAACTGGAGAACCATGCCCTGAAG 580
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
581 GTCTCTACATACCTGATGAGCAGATAACA...CAAGGTCTCAGAAATG 627
167 nProArgGlyArgArgGlyLeuGlnArgGlySerArgGlnGlyS 184
628 GCGTCGTGGA.....GGCTTTGGTCTCGGGGCCAGCCCGCAAGGGT 671
184 exPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
672 CGCCCTGGCAGCAGGGGCTCCAGCAAGCAGCAGCCAGTGGACATCCCT 721
198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysG 214
722 CTCCGGCTCTCTGGTGGCTACGCGAGTATGTAGGGCTATCATTTGGCA 771
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
772 GGTGGCCACCATCCGAACATCACAAACAGACGCGAGTCCAAATAGACG 821
231 aHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
822 TGCAATAGGAAGAGAAATCGGGCGCTCGGAGAAAGCCATCAGCGTGCAT 871
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMe 264
872 TCACCCCTGAAGCTGCTCTCCCGCTGCAAGATGATCTTGGAGATTAT 921
264 tHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
922 GCACAAAGGAGGCAAGGACACCAACAAACGCGAGATGAAGTTCCCTCA 971
281 leuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
972 TCTGGCTCATACAACTTCGTGGCGCACTCATTTGGCAAGGAGCGCG 1021
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
1022 AACCTGAAGAGTGGAGCAGACACAGAGACAGATCACCATCCTCATC 1071
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331

```

```

1072 GCTCCAGGACCTCACGCTCTATAACCTTGAGAGACCATCACGTGAAG 1121
331 lyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIle 347
1122 GCGCCAATTGAGAACTGTTGAGGGCCGAGGAGGATCATGAAGAAGTT 1171
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
1172 CGAGAGGCTTACGAGAACGACGTGGCGCCATGAGTTGCGAGTCCACCT 1221
364 uIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerG 381
1222 CATCCCTGGGCTTAACCTGCTGCTAGTCTCTCCAGCTTCATCCA 1271
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
1272 GCGCTGCTCCCTCT.....CTCCCAAGCAGTGTCACTGGGGCTGCT 1312
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
1313 CCCTATAGCTCTTCATGCAAGTCCGAGCAGGAGATGTTACAAGTGT 1362
411 eIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisI 428
1363 CATCCCGCGCCAGGCTGTGGCGCCATCATTTGCAAGAGGCGCCACCA 1412
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
1413 TCACCAACTCTCCCGTTTCGCCAGCGCTCCATCAAGATTGCTCCACCA 1462
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProGl 461
1463 GAAACACCTGACTCCCAAGTTCGAATGGTCTCATCTACTGGACCCCCAGA 1512
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGluA 478
1513 GGTCTAGTTCAAGGCTCAGGAAAGAAATTTATGCAAACTAAAAAGAAGA 1562
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
1563 ATTCTTTGGTCCCAAGGAGAAAGTAAAGTAGAGCCACATACGGGTT 1612
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThrValas 511
1613 CCGGCTTCAGCAGCCGCCGCTCATCGCAAAAGCGGCAAAACGGTGAA 1662
511 nGluLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnT 528
1663 TGAGCTGCAGAACTTACTGCTGAGGTGGTAGTGCCCAAGAGACCCAGA 1712
528 hrProAspGluAsnAspGlnValValLysIleThrGlyHisPheTyr 544
1713 CCCCGATGAGAACGACCAAGTCAATGTTAAGATCATCGGACATTTCTAT 1762
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnVally 561
1763 GCACGCGAGATGGCTCAGCGGAGATCCGAGACATCTCGCTCAAGTTAA 1812
561 sGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgA 578
1813 GCAA...CAGCACCAAGGGA...CAGAGCAACCTGGCCCGCAGGACGGA 1856
578 rglys 579
1857 GGAAG 1861

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-6

```

seq_documentation_block:
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng

```

APPLICANT: Gure, Ali
APPLICANT: Tsang, Solam
APPLICANT: Stockert, Elisabeth
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alexander
APPLICANT: Old, Lloyd J.
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
FILE REFERENCE: LUD 5538
CURRENT APPLICATION NUMBER: US/09/061,709B
CURRENT FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 8
SEQ ID NO 6
LENGTH: 3412
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-061-709-6

alignment_scores:
Quality: 1950.50 Length: 591
Ratio: 3.817 Gaps: 8
Percent Similarity: 86.464 Percent Identity: 65.313
alignment_block:
US-09-685-696-176 x US-09-061-709-6
Align seg 1/1 to: US-09-061-709-6 from: 1 to: 3412
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
73 ATGAACAAGCTTTATCATCGGAACCTTGAGCCCGCGCTCACCGCGGACGA 122
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
123 CCTCGGAGCTCTTTGGGACAGGAGCTGCCCTGGCGGACAGGTCC 172
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
173 TGCTGAAGTCGGGTACGCCCTTCGTGGACTACCCGACCAAGTGGGCC 222
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
223 ATCCGCGCATCGAGACCTCTCGGTAAAGTGAATTCATGGGAAAT 272
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
273 CATGGAAGTTGATTACTACGTCTCTAAAGCTAAGGAGCAGGAAATTC 322
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
323 AGATTGAAACATCCCTCCTCACCTGAGTGGAGGTGTTGGATGGACTT 372
101 LeuValGlnTyrGlyValValIleCysCysGluGlnValAsnThrAspSe 117
373 TTGGCTCAATATGGAGAGTGGAGAATGTGGAACAAGTCAACACAGACAC 422
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
423 AGAAACCCCTTGTCAACGTACATATGCAACAAGAGAGAAGCAAAAA 472
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
473 TAGCCATGGAGAAGTAAGCGGCATCAGTTTGAGAACTACTCTCTCAAG 522
151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnG 167
523 ATTTCCATACATCCCGGATGAAGAGGTGAGTCCCTTCGCCCCCTCAG.. 570
167 nProArgGlyArgGlyLeuGluGlnArgGlySerSerArgGlnGly. 183
571CGAGCCAGCGT.....GGGGACCACCTCTTCCCGGAGCAAGGCC 610

184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
611 ACGCCCTGGGGCAGCTTCTCAGGCCAGACAGATTGATTCGCCGTGCGG 660
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
661 ATCTCTGGTCCCAACCACTTTGTTGGTGCATCATCGGAAAGAGGGCTT 710
216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
711 GACCATAAAGAACATCATCAAGCAGACCCAGTCCCGGTAGATATCCATA 760
233 rgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
761 GAAAAGAGAACTCTGGAGCTGCAGAGAACCTGTCCACCATCATGCCACC 810
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMethIle 266
811 CCAGAGGGGACTTCTGAAGCATGCCGATGATCTTGAATCATGCAGAA 860
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
861 AGGCGAGATGAGCAAACTAGCGGAAGAGATTCCTCTGAAATCTTTGG 910
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
911 CACACAATGGCTTGGTGGAAAGACTGATTGGAAGAGAGGAGCAAAATTTG 960
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG 316
961 AAGAAAATTTGAACATGAACAGGAGCAAGATAACAATCTCATCTTTGCA 1010
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
1011 GGATTGAGCATATACACCCGGAAGAACCATCACTGTGAAGGCGACAG 1060
333 alGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgGlu 349
1061 TTGAGGCTGTGCCAGTCTGAGATAGAGATTATGAAGAAGCTGCGTGAG 1110
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1111 GCCTTTGAAAATGATGCTGGCTGTTAAACCAACAGCAATCTGATCCC 1160
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe..... 376
1161 AGGTTGAACTCAGGCGACTTGGCATCTTTCAACAGGACTGTCCGTGC 1210
377ProProThrSer.....GlyMetProProProThr..... 386
1211 TATCTCCACAGCAGGCGCCCGCGAGCTCCCGCGTCCCGCTTACCAC 1260
387SerGlyProProSerAlaMetThrPro..... 395
1261 CCCTTCACTACCCACTCCGGATCTTCTCCAGCTGTATACCCCATCACA 1310
396ProTyrProGlnPheGluGln...SerGluThrGluThrValH 409
1311 GTTTGGCCCGTTCGGCATCATCACTCTTATCCAGAGGAGGATTTGCA 1360
409 isGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425
1361 ATCTCTTCATCCCAACCCAGGCTGTGGCGGCATCATCGGAAAGAGGG 1410
426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAl 442
1411 GCACACATCAAAACAGCTGGGAGATTCGCGGAGCCTCTATCAAGATTGC 1460
442 aProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyP 459
1461 CCCTGCGGAAGGCCACAGCTCAGCGGAAGGATGGTCAATCATCACCAGCC 1510


```
459 roproGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLys 475
1511 CACCGGAAGCCCAAGTTCAAGGCCAGGGACGGATCTTTGGGAACATAA 1560
476 GluGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIle 492
1561 GAGGAAAACCTCTTAAACCCCAAGAAAGTGAAGCTGGAAGCGCATAT 1610
492 eArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysT 509
1611 CAGAGTGCCCTCTCCACAGCTGGCGGGTGATTGGCAAAAGTGGAAGA 1660
509 hrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArg 525
1661 CCGTGAAACCAAGCTGCAAGATTTAAACAGTGCAGAGTCACTGCGCTCGT 1710
526 AspGlnThrProAspGluAsnAspGlnValValLysIleThrGlyHis 542
1711 GACCAACACCCAGATGAAATGAGGAAGTGTGTCAGAAATATATCGGGCA 1760
542 sPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrG 559
1761 CTTCTTGTGCTACCCAGACTGCACAGCGCAAGATCAGGGAATTTGTACAAC 1810
559 InValLysGlnHisGlnGlnGln 566
1811 AGGTGAAGCAGCAGGAGCAGAAA 1833
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-061-709-7

```
seq_documentation_block:
; Sequence 7, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tsung
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; FILE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061.709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-7
```

alignment_scores:
Quality: 1873.00 Length: 509
Ratio: 4.116 Gaps: 9
Percent Similarity: 89.391 Percent Identity: 74.656

alignment_block:

US-09-685-696-176 x US-09-061-709-7 ..

Align seg 1/1 to: US-09-061-709-7 from: 1 to: 1946

```
77 ArgGlnArgIleArgLysLeuGlnIleArgAsnIleProHisLeuG1 93
102 CGGGTTAGGAGCGGAAAAATCCAAATCCGAAATATTCACCCCGAGTCCG 151
93 nTrpGluValLeuAspSerLeuLeuValGlnTyrGlyValValGluSerC 110
152 ATGGGAAGTACTGGACAGGCTGCTGGCTCAGTATGTCAGTAGAGAACT 201
```

```
110 ysGluGlnValAsnThrAspSerGluThrAlaValValAsnValThrTyr 126
202 GTGAGCAAGTGAACACCGAGTGAAGCGCAGTGTGAATGTCACTAT 251
127 SerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheG1 143
252 TCAACCGGGAGCAGACCGCAAGCATCATGAAGCTGAATGGCCACCA 301
143 nLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGluMetAla 160
302 GTTGGAGAACCATGCCCTGAAGGCTCTTACATCCCGGATGAGCAGATAG 351
160 laGlnGlnAsnProLeuGlnGlnProArgGlyArgGlyLeuGlyGln 176
352 CA...CAGGACCTGAGAATGGCGCGCGAGGG.....GGCTTTGGCTCT 392
177 ArgGlySerSerArgGlnGlySerPro.....GlySerValSerLys 190
393 CGGGGTGACGCCCGCGCGGCTCACCTGTGGCAGCGGGGGCCCGACCAA 442
190 sGlnLysProCysAspLeuProLeuArgLeuValProThrGlnPheV 207
443 GCAGCAGCAAGTGGACATCCCCCTTCGGCTCTGGTGGCCACCCAGTATG 492
207 alGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLys 223
493 TGGGTGCCATTTATTGGCAAGAGGGGGCCACCATCCGCAACATCACAAAA 542
224 GlnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaA1 240
543 CAGACCCAGTCCAAGATAGACGTGCATAGGAGGAGAACCGAGTGCAGC 592
240 aGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaC 257
593 TGAAGGAGCATCTGTGGAGATTATGCATAAGAGGCTAAGGACACCAAAACG 692
257 ysLysSerIleLeuGluIleMethHisLysGluAlaGlnAspIleLysPhe 273
643 GTAAGATGATCTTGGAGATTATGCATAAGAGGCTAAGGACACCAAAACG 692
274 ThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyAr 290
693 GCTGACGAGGTTCCTCCCTGAGATCTTCGCCCATATAAATTTGTAGGCG 742
290 gLeuIleGlyLysGluGlyArgAsnLeuLysIleGluGlnAspThrA 307
743 TCYCATTTGCCAAGAGGACGGAACTTGAAGAGGTAGACCAAGATACCG 792
307 spThrLysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnPro 323
793 AGACAAAATACCATCTCTCTGTTGCAAGACCTTACCCCTTTACAACCT 842
324 GluArgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaG1 340
843 GAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTTGTCAGGGCCGA 892
340 uGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlas 357
893 GCAGGAATAATAAGAAAGGCTCGGAGGCTATGAGAATGTATGGCTG 942
357 erMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeu 373
943 CCATGAGC.....TCTACCTGATCCCTGGCCTGAACCTGGCTGCTGTA 986
374 GlyLeuPheProProThrSerGlyMetProProProThrSerGlyProPr 390
987 GGTCTTTTCCAGCTTCATCCAGCGAGTCCCGC.....CCTCC 1027
390 oSerAlaMetThr.....ProProThrProGlnPheGluGlnSer....G 404
1028 CAGCAGGTTACTGGGGCTGCTCCCTATAGTCTCTTTATCGAGCTCCCG 1077
404 luThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIle 420
```

```
1078 AGCAGGATGGTGCAGGTGTTTATCCCGCCGAGTGGCGCCATC 1127
421 lIeGlyLysGlnGlyGlnHisLysGlnLeuSerArgPheAlaGlyAl 437
1128 ATCGCAAGAGGGGAGCAGCATCAACAGCTCTCCCGTTTGCAGGCG 1177
437 eSerLysLysLysAlaProAlaGluAlaProAspAlaLysValArgMetV 454
1178 CTCCATCAAGATGTCACCCAGCCGAAACACCTGACTCCAAAGTTCGTATGG 1227
454 alIleIleThrGlyProGluAlaGlnPheLysAlaGlnGlyArgIle 470
1228 TTATCATCACTGGACCGCCAGAGCCCAATTCAGGCTCAGGGAGAAATC 1277
471 TyrGlyLysLysLysGluGluAsnPheValSerProLysGluGluVally 487
1278 TATGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTAA 1327
487 sLeuGluAlaHisLysArgValProSerPheAlaAlaGlyArgValIleG 504
1328 GCTGGAGACCCACATACGTGTGCAGCATCAGCAGCTGGCGGTCATATG 1377
504 lYlYsGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGlu 520
1378 GCAAGGTGGAACCGTGAACAGTTCAGAGATTTGACGGCAGCTGAG 1427
521 ValValValProArgAspGlnThrProAspGluAsnAspGlnValValVa 537
1428 GTGGTAGTACCAAGAGACAGACCCCTGTATGAGAACGACCAAGGTATCGT 1477
537 lYsIleThrGlyHisPheThrAlaCysGlnValAlaGlnArgLysIleG 554
1478 GAAATCATCGGCATTTCTATGCGCATCAGATGGCTCAACGGGAATGCC 1527
554 InGluIleLeuThrGlnValLysGlnHisGlnGlnLysAlaLeuGln 570
1528 GAGCATCCTGGCCCGAGGTAAAGCAG...CAGCATCAGAAAGGGA...CAG 1571
571 SerGlyProProGlnSerArgLys 579
1572 AGTAACCAAGGCCAGGACGGAGGAAG 1598
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-061-709-8

```
seq_documentation_block:
; Sequence 8, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Kouth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-8
```

alignment_scores:
Quality: 1868.00
Ratio: 3.820
Length: 568
Gaps: 6

```
Percent Similarity: 86.092 Percent Identity: 64.789
alignment_block:
US-09-685-696-176 x US-09-061-709-8
Align seg 1/1 to: US-09-061-709-8 from: 1 to: 3283
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
73 ATGAACAAGCTTTTACATCGGAACCTTGAGCCCGCGCTCACCGCGGACGA 122
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyPropheL 34
123 CCTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCTTGGCGGACAGGPTCC 172
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTirpAla 50
173 TGCTGAAGTCGGGCTACGCCCTTCGTGGACTACCCCGACCAAGACTGGGCC 222
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
223 ATCCGCGCCATCGAGACCTCTCGGTTAAAGTGAATTCATGGGAAAAT 272
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
273 CATGGAAGTTGATTACTCAGTCTCTAAAAGACTAAGGAGCAGGAAAAATTC 322
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
323 AGATTCCGAACATCCCTCCTCCTCAGCTGAGTGGGAGGTGTGGATGGACTT 372
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
373 TTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAC 422
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
423 AGAAACCGCCGTTGTCAACGCTCAGATATGCAACAAGAGGAAGCAAAAA 472
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
473 TAGCCATGGAGAGCTAAGCGGCGCATCAGTTTGAGAACTACTCTCTCAAG 522
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
523 ATTTCTTACATCCCGATGAAGAGTGAAGTCCCTTCGCCCTCAG... 570
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly. 183
571 ....CGAGCCCGAGCGT.....GGGACCACTCTTCCCGGAGCAAGGCC 610
184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
611 ACGCCCTGGGGCAGCTCTCAGCCAGAGAGAGTGTGATTTCCCGCTCGG 660
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlyAl 216
661 ATCCTGGTCCCCACCGATTTGTTGGTGCCATCATCGAAAGGAGGCGTT 710
216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 233
711 GACCATAAAGAACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATA 760
233 rGlyGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
761 GAAAAGAGAACTCTGGAGCTGAGAGAGAGCTGTACCATTCATGCCACC 810
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLy 266
811 CCAGAGGGGACTTCTGAAGCATGCCGATGATTTCTTCAAAATCATGCAGA 860
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
```

```

seq_name: /cn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-5

seq_documentation_block:
; Sequence 5, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 5
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-5

alignment_scores:
Quality: 1637.00 Length: 453
Ratio: 4.093 Gaps: 9
Percent Similarity: 88.300 Percent Identity: 74.172

alignment_block:
US-09-685-696-176 x US-09-061-709-5 ..
Align seg 1/1 to: US-09-061-709-5 from: 1 to: 1708

133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGlnAsnPheThrLe 149
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
32 CGGGAGGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCGCT 81
149 uLysValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuG 166
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 GAAGTCTCTCTACATCCCCCATGAGCATAGCA...CAGGACCTGAGA 128
166 lnGlnProArgGlyArgGlyLeuGlnArgGlySerSerArgGln 182
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
129 ATGGCGCCGAGGG.....GGCTTTGGCTCTCGGGCTCAGCCCGCCAG 172
183 GlySerPro.....GlySerValSerLysGlnLysProCysAspLe 196
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
173 GGCTTCACCTGTGGCGGGGGGGCCCAAGCAGCAGCAAGTGGAGAT 222
196 uProLeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyL 213
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 CCCCCTTCGGCTCTGTGGTCCCAACCATGATGTGGGTGCCATATTGGCA 272
213 ysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIle 229
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 AGGAGGGGGCCACCATCCGCAACATCAAAACAGACCCAGTCCAAGATA 322
230 AspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrI 246
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 GACGTGATAGGAAGAGAGACGCGAGGTGCAGCTGAAAAGCCATCACTGT 372
246 eLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluI 263
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 GCATCTCCACCCCTCAGGGCTCTCTCTCGCTTGTAAAGATGATCTGGAGA 422
263 leMetHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeu 279
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
423 TTATGCAATAAGAGGCTTAAGGACACCAAAACGGCTGACGAGGTCCCTG 472

```

```
280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluCl 296
|||||
473 AAGATCCTGGCCCAATAAATACCTTTAGGGCTCTCATTTGGCAAGGAAG 522
|||||
296 YArgAsnLeuLysIleGlnAspThrAspThrLysIleThrIleS 313
|||||
523 ACGGAACCTGAAGAGGAGTAGAGCAAGATACCGAGACAAAATCACCATCT 572
|||||
313 erProLeuGlnLeuThrLeuTyAsnProGluArgThrIleThrVal 329
|||||
573 CCTCGTTGCAAGACCTTACCCTTTACAAACCTTGAGAGACCATCACTGTG 622
|||||
330 LysGlyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLy 346
|||||
623 AAGGGGCCATCGAGATTTGTCAGGGCCGAGCAGGAGAAATAATGAAGA 672
|||||
346 sIleArgGluSerTyGluAsnAspIleAlaSerMetAsnLeuGlnAlaH 363
|||||
673 AGTCGGGAGGCCTATGAGATGATGTGCTGCCATGAGC.....TCTC 716
|||||
363 IsLeuIleProGlyLeuAsnLeuAlaLeuGlyLeuPheProProThr 379
|||||
717 ACCTGATCCTGGCTGAACCTGGCTGCTGTAGTCTTTTCCACCTTCA 766
|||||
380 SerGlyMetProProThrSerGlyProProSerAlaMetThr..... 394
|||||
767 TCCAGCGCAGTCCGCCG.....CCTCCAGCAGCGTTACTGGGCG 807
|||||
395 .ProProTyProGlnPheGluGlnSer...GluThrGluThrValHisG 410
|||||
808 TGCTCCCTATAGCTCTTTATGCAAGTTCCTCCGAGCAGGATGTGCAG 857
|||||
410 InPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGln 426
|||||
858 TGTATTATCCCGCCCGCAGCAGTGGCGCCATCATCGGCAAGAGGGCGAG 907
|||||
427 HisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaPr 443
|||||
908 CACATCAACAGCTCTCCGGTTTGGCCAGCGCTCCCATCAAGATTGCACC 957
|||||
443 oAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyProp 460
|||||
958 ACCGAAACACTGACTCTCAAGTTCGTATGGTTATCATCTACTGGACCGC 1007
|||||
460 roGluAlaGlnPheLysAlaGlnGlyArgIleTyGlyLysIleLysGlu 476
|||||
1008 CAGAGGCCCAATTCAAGCTCAGGAGAGATCTATGGCAAACTCAAGGAG 1057
|||||
477 GluAsnPheValSerProLysGluValLysLeuGluAlaHisIleAr 493
|||||
1058 CAGAACTTCTTTGGTCCCAAGGAGGAGTGAAGCTGGAGACCCACATAGC 1107
|||||
493 gValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThrV 510
|||||
1108 TGTCCAGCATCAGCAGCTGGCCGGGTCTTGGCAAAAGTGGAAAAACGG 1157
|||||
510 aAsnGluLeuGlnAsnLeuSerAlaGluValValValProArgAsp 526
|||||
1158 TGAACGAGTTCCAGATTTGACGGCAGCTGAGTGGTAGTACCAGAGAC 1207
|||||
527 GlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHisPh 543
|||||
1208 CAGACCCCTGTATGAGACGACCGACGCTATCGTGAAATCATCGGACATTT 1257
|||||
543 eTyAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnV 560
|||||
1258 CTATGCCAGTCATGAGTGTCAACGAGATCCGAGACATCTCTGGCCCGAGG 1307
|||||
560 aLysGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSer 576
|||||
1308 TTAAGCAG...CAGCATCAGAAGGA...CAGAGTAACCGGCCCGCAGCA 1351
|||||
```

```
577 ArgArgLys 579
|||||
1352 CGAGGAAG 1360

seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-08-021-608D-1

seq_documentation_block:
; Sequence 1, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2384
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 473 bp
; OTHER INFORMATION: variable region where R is A or G.
US-08-021-608D-1

alignment_scores:
Quality: 241.00 Length: 449
Ratio: 1.076 Gaps: 17
Percent Similarity: 49.889 Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-021-608D-1
..
Align seg 1/1 to: US-08-021-608D-1 from: 1 to: 2384
161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgArgGlyLe 174
|||||
213 CAAAAAGACCTTTAGAGATGGAGATCAACCATGCTTAAGAAAGTTGC 262
```

```
174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
263 TCCTCAAAATGACTCTTTTGGACACAGATTACACCGATGCATCAGCAGC 312
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
191 lnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
313 AAAGCAGATCTGTAATGACAGAGAATAACAAAGTTCCAGATGGAATGGTT 362
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
208 GlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
363 GGATTCAATATTCGACAGAGGTGAACAGATCTCAGCATACAAACAGGA 412
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaG 241
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
413 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGGTGGCTTCCAG 459
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
241 lnLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
460 AAAGGTCTCTGATRTTAACCTGGAAACACCTGAATCTGTCCAGTCAGCAAAA 509
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
510 CGGTTACTGGACCATGATTGTTGAAAAGGAAGACACCTCCTGGCTTCCA 559
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
560 TCATGGCGATGGACCGGAATACAGTTCAGAAATCATGATTCACGCTA 609
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
610 GCAAGCGCAGGATTAGTCATTTGGAAGGAGGAGGAGAACTATTAAACAGCTT 659
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
660 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 698
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
319 rLeuTyAsnProGluArgThr.....IleThrValLysG 331
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
699 .....GGGCCGACAGAACACTGGTGTGACAAACCTCTTAGGATTACAG 741
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
331 lysnValGluThrCysAlaLysAlaGluGluLeuMetLysLysIle 347
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
742 GAGACCCATATAAAGTTCAACAGCAAGCAAGGAAATGGTTAGATTAAAT 791
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
348 ArgGlu.....SerTyGluAsnAspIleAlaSerMetas 359
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
792 CGTGATCAAGCGGTTTCAGAGAAGTTCCGAATGAGTATGGTCAAGA... 839
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaGlyLeuP 376
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
840 .....ATAGGAGGA..... 848
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
376 heProThrSerGlyMetProProProThrSerGlyProProSerAla 392
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
849 .....AATGAAGGATAGATGTCCCC..... 869
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
393 MetThrProProTyProGlnPheGluGlnSerGluThrGluThrValHi 409
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
869 ..... 869
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyG 426
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
870 .....ATCCAGAGATTCTGTGGCATGTGTAATAGGAAGAAATCGAG 912
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
426 lnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
913 AGATGATCAAAAAAATACAAATGATGCTGGTGTTCGCAATTCAGTTTAAAG 962
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
963 CCAGATGATGGCAACACCCGAA...AGGATAGCACAAAATAACAGGACC 1009
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
```

```

; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 473 bp
; OTHER INFORMATION: variable region where R is A or G.
US-08-726-160-1

alignment_scores:
  Quality: 241.00      Length: 449
  Ratio: 1.076        Gaps: 17
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-726-160-1 ..
Align seg 1/1 to: US-08-726-160-1 from: 1 to: 2384

161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgArgGlyLe 174
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 CAAAAAGACCTTTAGAGATGGAGATCAACACAGATGCTAAGAAAGTTGC 262
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
263 TCCTCAAAATGACTCTTTTGGAAACACAGATTACCCCGCATGCATCAGCAGC 312
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

191 lnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
313 AAGCAGATCTGTAATGACAAAGAAATCAAAAGTTCCAGATGGATGGTT 362
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

208 GlyAlaIleLeuGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
363 GGATTCAATATGGCAGAGGAGGTGAACAGATCTCACGCATACACAGGA 412
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaLag 241
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
413 ATCTGGATGCAAAATACAGATA...GCTCTGACAGTGGTGGCTTCCAG 459
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

241 luLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
460 AAGGTCTCTGTATTTAACTGGAACACCTGTAATCTGTCACATCAGCAAAA 509
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
510 CGGTTACTGGACAGATTGTTGAAAAAGGAAGACAGCTCTCTGCTTCCA 559
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
560 TCATGGCGATGGCGGAAATCAGTTCAAGAAATCATGATTCACAGCTA 609
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
610 GCAAGCAGGATTAGTATTGGAAGAGGGGAGAACTATTAAACAGCTT 659
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
660 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 698
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

319 rLeuTyAsnProGluArgThr.....IleThrValLysG 331
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
699 .....GGGCCGACAGACACTGTGTGTGCAAAACCTCTTAGGATTACAG 741
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
742 GAGACCCATATAAGTTCAACACCCAGGAAATGGTTAGATTAAATT 791
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetAs 359
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
792 CQTGATCAAGCGGTTTCAGAGAAAGTTCCGAATGAGTATGGTCAAGA... 839
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuP 376
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
840 .....ATAGGAGGA..... 848
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

376 heProProThrSerGlyMetProProThrSerGlyProProSerAla 392
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
849 .....AATGAGGGATAGATGTCGCC..... 869
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

393 MetThrProProTyrProGlnPheGluInSerGluThrGluThrValHi 409
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
869 ..... 869
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyG 426
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 .....ATTCCAAGATTGCTGTTGGCATTGTAATAGGAAGAAATGGAG 912
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

426 lnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
913 AGATCATCAAAAAAATACAAATGATGCTGCTTCGCATTTCAGTTTAAAG 962
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

443 ProLaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
963 CCAGATGATGGACAAACACCCGAA...AGGATAGCACAAATAACAGGACC 1009
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

459 oProGlu...AlaGlnPheLysAlaGln..... 467
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1010 TCCAGACCGATGTCAACATGCTGCAGAAATATTATACAGACCTTCTTCGAA 1059
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
468 .....GlyArgIleTyr 471
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

1060 GTGTTTCAGCTGGTAATCTCTGTGGACCTGGACCTGGTGGTCGAGGAAGA 1109
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 GlyLys.....IleLysG 476
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1110 GGTAGAGGTCAAGGCAACTGGAACATGGGACCACCTGGTGGATTACAGGA 1159
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

476 uGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIleA 493
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1160 ATTAATTTTATT..... 1172
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1173 ..GTGCCAACTGGGAAACTGGATTAAATAGGAAAGAGGTGAACACC 1220
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

510 ValAsnGluLeuGlnAsnLeuSerSerAlaGluValValValProArgAs 526
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1221 ATAAAAGCATAGCCAGCAGCTCTGGTGCAGAAATAGAACTTCAGAGAAA 1270
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1271 TCCTCCCAAAATGCAGATCCTAATATGAAGTTATTACAATTCGTGGCA 1320
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

542 is.....PheTyrAlaCysGlnValAlaGlnArgLysIle 553
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1371 CTCACACACATACACTATGCTCGCAACTCATAGAGAAAAGATT 1367
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: /cgn2.6/ptodata/2/ina/PTUS_COMB.seq:PCT-US94-01782-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN

```

STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/01782
 FILING DATE: 22-FEB-1994
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. 08/021,608
 FILING DATE: 22-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: WILLIAM S. FEILER
 REGISTRATION NUMBER: 26,728
 REFERENCE/DOCKET NUMBER: 2026-4063PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2384
 TYPE: Nucleic acid
 STRANDEDNESS: Double
 TOPOLOGY: Unknown
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: No
 ANTI-SENSE: YES
 ORIGINAL SOURCE:
 ORGANISM: Human

STRAIN:
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE:
 HAPLOTYPE:
 TISSUE TYPE:
 CELL TYPE:
 CELL LINE: HL60
 ORGANELLE:
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT:
 MAP POSITION:
 UNITS:
 FEATURE:

NAME/KEY:
 LOCATION:
 IDENTIFICATION METHOD:
 OTHER INFORMATION: 473 bp
 OTHER INFORMATION: variable region where R is A or G.
 PCT-US94-01782-1

alignment_scores:
 Quality: 241.00 Length: 449
 Ratio: 1.076 Gaps: 17
 Percent Similarity: 49.889 Percent Identity: 22.272

alignment_block:

US-09-685-696-176 x PCT-US94-01782-1 ..

Align seg 1/1 to: PCT-US94-01782-1 from: 1 to: 2384

161 GlnProArgGlyArgGlyLeu.....GlnProArgGlyArgGlyLeu 174
 |||.....|
 213 CAAAAAGACCTTTAGAGATGGAGATCAACCGATGCTTAAGAAAGTTGC 262
 |||.....|
 174 uGlyGlnArgGlySerArgGlnGlySerProGlySerValSerLysG 191
 |||.....|
 263 TCCTCAAAATGACTCTTTTGGAAACAGATTACCACCGATGCATCAGCAGC 312

191 InLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
 |||.....|
 313 AAAGCAGATCTGTAAATGACAGAAGATAACAAGTTCAGATGAATGGTT 362
 |||.....|
 208 GlyAlaIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
 |||.....|
 363 GGATTTCATAATGGCAGAGGAGTGAACAGATCTCAGCATACACAGGA 412
 |||.....|
 224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaG 241
 |||.....|
 413 ATCTGGATGCAAAATACAGATA...GCTCTGACAGTGGTGGCCTTCCAG 459
 |||.....|
 241 LysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
 |||.....|
 460 AAAGTCTCTGTATRTTAACCTGGAACACCTCAATCTCTCCAGTCAGCAAA 509
 |||.....|
 258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
 |||.....|
 510 CGGTTACTGGACACAGATTGTTGAAAAAGGAAGACACAGCTCCTGGCTTCCA 559
 |||.....|
 274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
 |||.....|
 560 TCATGGCGATGGACCGGAAATGCAGTTCAAGAAATCATGATTCCAGCTA 609
 |||.....|
 286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysIle 302
 |||.....|
 610 GCAAGGCAGGATTAGTCATTGGAAAAGGGGAGAACTATTAAACAGCTT 659
 |||.....|
 303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
 |||.....|
 660 CAGGAACGGGCTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 698
 |||.....|
 319 rLeuTyrAsnProGluArgThr.....IleThrValLysG 331
 |||.....|
 699GGGCCGACAGACACTGGTCTGCACAAACCTCTTAGGATTACAG 741
 |||.....|
 331 LysAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
 |||.....|
 742 GAGACCCATATAAAGTTCACCAAGCCCAAGGAATGGTGTAGAGTTAATT 791
 |||.....|
 348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetAs 359
 |||.....|
 792 CDTGATCAAGCGGTTTCAGAGAAGTTCCGAATGAGTATGGGTCAAGA... 839
 |||.....|
 359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuP 376
 |||.....|
 840ATAGGAGGA..... 848
 |||.....|
 376 heProThrSerGlyMetProProThrSerGlyProProSerAla 392
 |||.....|
 849AATGAAGGATAGATGTCCCC..... 869
 |||.....|
 393 MetThrProProTyrProGlnPheGluGlnSerGluThrGluThrValH 409
 |||.....|
 869 869
 |||.....|
 409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyG 426
 |||.....|
 870ATTCCAAGATTTCGTTGGCATTGTAATAGGAAGAAATGGAG 912
 |||.....|
 426 InHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
 |||.....|
 913 AGATGATCAAAAAAATACAAATGATGCTGGTGTTCGCATTCAGTTTAAG 962
 |||.....|
 443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
 |||.....|
 963 CCAGATGATGGGACACACCCGAA...AGGATAGCAAAATAACAGCACC 1009
 |||.....|
 459 oProGlu...AlaGlnPheLysAlaGln..... 467
 |||.....|
 1010 TCACAGACCGATGTCACATGCTGCAGAAATATTATACAGACCTTCTTCGAA 1059
 |||.....|

```
468 .....GlyArgIleTyr 471
1060 GTGTTCAGCGTGGTAATCTGCTGGACCTGGACCTGGTGGTCGAGAGA 1109
472 GlyLys.....IleLysG1 476
1110 GGTAGAGTCAAGCACTGGAACATGGACCACCTGGTGATTACAGGA 1159
476 uGluAsnPheValSerProLysGluGluValLysLeuAlaHisIleA 493
1160 ATTTAATTTATT..... 1172
493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
1173 ..GTGCCAACTGGGAAACTGGATTAATAATAGGAAAGGAGGTCAAACC 1220
510 ValAsnGluLeuGluAsnLeuSerSerAlaGluValValProArgAs 526
1221 ATAAAAGCATNAGCCAGCAGCTCTGGTGCAGAATAAGACTTCAGAGAA 1270
526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGly 542
1271 TCCCTCCACCAATGCAGATCTAATATAGAGTTATTACATTCGTGCA 1320
542 is.....PheTyrAlaCysGlnValAlaGlnArgLysIle 553
1321 CTCACACACAGATAGACTATCTCGGCAACTCATAGAGAAAGATT 1367
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-9
seq_documentation_block:
; Sequence 9, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: CDNA
```

```
; HYPOTHETICAL: No
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 470 bp variable
; OTHER INFORMATION: region where R is A or G.
US-08-021-608D-9

alignment_scores:
  Quality: 236.50      Length: 449
  Ratio: 1.056        Gaps: 18
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-021-608D-9 ..
Align seg 1/1 to: US-08-021-608D-9 from: 1 to: 2381
161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgGlyLe 174
174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
263 TCCTCAAAATGACTCTTTTGGACACACAGTTACCA...CCGATGCATCAGC 309
191 lnLysProCysAspLeuProLeuArgLeuValProThrGlnPheVal 207
310 AGCAAAGATCTGTAATGACAGAAGAATAACAAAGTTCCAGATGGAATGGTT 359
208 GlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG1 224
360 GGATTCATAATTGCAGGAGGTGACACAGATCTCACGCATACACACAGA 409
224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaIaG 241
410 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGGTGGCTTCCAG 456
241 luLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
457 AAAGTCTCTGTATTTAACTGGAACACCTGAATCTGTCCAGTCAGCAAAA 506
258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
507 CGGTACTGGACAGATGTTGAAAAGGAAGACACAGCTCCTCGCTTCCA 556
274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
557 TCATGGCGATGGACCGGAAATGCAGTTCAAGAAATCATGATTCAGCTA 606
286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
607 GCAAGGAGGATTAGTTCATTTGAAAAGGGGAGAACTATTAAACAGCTT 656
303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
657 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAGAC..... 695
319 rLeuTyrAsnProGluArgThr.....IleThrValLysG 331
696 .....GGGCCGAGAACACTGGTGTGCACAAACCTCTTAGGATTACAG 738
331 lyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIle 347
739 GAGACCCATATAAAGTTCAACAGCCAGGAATGGTGTAGAGTTAATT 788
348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetAs 359
789 CGTGATCAAGCGGCTTTCAGAGAGAAGTTCCGGAAATGAGTATGGTCAAGA... 836
```



```

359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuP 376
      ||| |||
837 .....ATAGGAGGA..... 845
376 heProProThrSerGlyMetProProProThrSerGlyProProSerAla 392
      : : : : :
846 .....AATGAAGGATAGATGTCGCC..... 866
393 MetThrProProTyProGlnPheGluGlnSerGluThrGluThrValHi 409
866 ..... 866
409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 426
      : : : : :
867 .....ATTCCAAGATTGCTGTGGCATTGTAAATAGGAAGAAATGGAG 909
426 InHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
      : : ||||| : : : : :
910 AGATGATCAAAAAATACAAAATGATGTGGTGTTCGCATTTCAGTTTAAG 959
443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
      ||| : : : : :
960 CCAGATGATGGACACACCCGAA...AGATAGCACAAATACAGACC 1006
459 oProGlu...AlaGlnPheLysAlaGln..... 467
      ||||| : : : : :
1007 TCCAGACCGATGTCAACATGCTGCAGAAATATTACAGACCTTCTTCGAA 1056
468 .....GlyArgIleTyr 471
      |||||
1057 GTGTTCAGCTGGTAAATCCTGTGTGGACCTGGACCTGGTCTGAGGAAGA 1106
472 GlyLys.....IleLysGI 476
      ||||| : : : : :
1107 GTGATGAGTCAGCAACTGGAAACATGGACACCTGGTGGATTACAGGA 1156
476 uGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIleA 493
      | |||||
1157 ATTTAATTTATT..... 1169
493 rgValProSerPheAlaGlyArgValIleGlyLysGlyLysThr 509
      ||||| : : : : :
1170 ..GTGCCAACTGGGAAACTGGATTAAATAGGAAAGAGGTGAACCC 1217
510 ValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArgAS 526
      : : : : :
1218 ATAAAAGCATAAGCCAGCAGTGTGTGGCAAGAAATAGAACTTCAGAGAAA 1267
526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
      : ||||| : : : : :
1268 TCCTCCACCAATGCAGACTCTTAATATGAATTATTACAAATTCGTGGCA 1317
542 Is.....PheTyrAlaCysGlnValAlaGlnArgLysIle 553
      ||||| : : : : :
1318 CTCCACACAGATAGACTATGTCTGCCAACTCATAGAAGAAAAGATT 1364

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq;us-08-726-160-9

seq_documentation_block:
; Sequence 9, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA

```

```

? ZIP: 10154
? COMPUTER READABLE FORM:
? MEDIUM TYPE: FLOPPY DISK
? COMPUTER: IBM PC COMPATIBLE
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: WORDPERFECT 5.1
? CURRENT APPLICATION DATA:
? FILING DATE: 04-OCT-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/021,608
? FILING DATE: 22-FEB-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: WILLIAM S. FEILER
? REGISTRATION NUMBER: 26,728
? REFERENCE/DOCKET NUMBER: 2026-4063US1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 758-4800
? TELEFAX: (212) 751-6849
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2391
? TYPE: Nucleic acid
? STRANDEDNESS: Double
? TOPOLOGY: Unknown
? MOLECULE TYPE: cDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: YES
? ORIGINAL SOURCE:
? ORGANISM: Human
? CELL LINE: HL60
? FEATURE:
? OTHER INFORMATION: 470 bp variable
? OTHER INFORMATION: region where R is A or G.
US-08-726-160-9

alignment_scores:
    Quality: 236.50      Length: 449
    Ratio: 1.056         Gaps: 18
Percent Similarity: 49.889      Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-726-160-9      ..

Align seg 1/1 to: US-08-726-160-9      from: 1 to: 2381

161  GlnGlnAsnProLeuGln.....GlnProArgGlyArgArgGlyLe 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213  CAAAAGAAGACCTTTAGAAGATGGAGATCAACCATGCTTAAGAAAGTTGC 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174  uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263  TCCTCAAAATGACTCTTTTGGAAACACACATTACCA...CCGATGCATCAG 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191  lnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310  AGCAAAAGATCTGTAATGACAGAGAAGATCAACAAGTTCACAGATGGATGT 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208  GlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGl 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360  GGATTCAATAATTGCAGAGGAGGTGAACAGATCTCACGCATACACAGGA 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224  nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaAg 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410  ATCTGGATGCAAAATACAGATA...GCTCCTGCACAGTGGTGCCTCCAG 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241  luLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
457  AAAGGTCCTATATTTAACTGGAACACCTGTAATCTGTCCAGTCAGCAAAA 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 470 bp variable
; OTHER INFORMATION: region where R is A or G.
PCT-US94-01782-9

alignment_scores:
  Quality: 236.50      Length: 449
  Ratio: 1.056         Gaps: 18
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x PCT-US94-01782-9
Align seg 1/1 to: PCT-US94-01782-9 from: 1 to: 2381

161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgGlyLe 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 CAAAAAGACCTTTAGAGATGGAGATCAACCCAGATGCTAAGAAAGTTGC 262

174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 TCCTCAAAATGACTCTTTTGGAAACAGTTACCA...CGATGTCATCAGC 309

191 lnLysProCysAspLeuProLeuArgLeuValProThrGlnPheVal 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 AGCAAGATCTGTATGACAGAGAGATACAAAGTTCCAGATGGATGGTT 359

208 GlyAlaIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 GGATTTCATAATGGCAGAGAGGTGAACAGATCTCAGCGCATACAACAGGA 409

224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAag 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGTGGCCTTCCAG 456

241 luLysSerIleThrLeuSerThrProGluGlyThrSerAlaAlaCys 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
457 AAAGTCCTGTATRTAACTGGACACCTGAATCTGTCAGTCAGCAAAA 506

258 LysSerIleLeuGluIleMethHisLysGluAlaGlnAspIleLysPhe 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
507 CGGTTACTGGACAGATTTGTAAGAAAGGAGACACAGCTCCTGGCTTCCA 556

274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 TCATGGCGATGGACCGGAAATGCAAGTTCAGAAATCATGATTCCAGCTA 606

286 snPheValGlyArgLysGlyLysGluGlyArgAsnLeuLysLysIle 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 GCAAGCGAGATGATGTCATGGAAGAGGGGAGAACTATTAAACAGCTT 656

303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 695

319 rLeuTyAsnProGluArgThr.....IleThrValLysG 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
696 .....GGGCGCGAGAACACTGGTCTGACAAACCTCTTAGGATTACAG 738

331 lYasnValGluThrCysAlaLysAlaGluGluIleMethLysLysIle 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
739 GAGACCCATATAAGTTCAACAGCCCAAGAAATGGTGTAGAGTTAAT 788

348 ArgGlu.....SerTyGluAsnAspIleAlaSerMetAs 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
789 CGTGATCAAGCGGCTTTCAGAGAAAGTTCCGAATGAGTATGGGTCAAGA.. 836

359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyLeup 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
837 .....ATAGGAGGA..... 845

```

```

376 heProProThrSerGlyMetProProProThrSerGlyProProSerAla 392
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 .....AATGAAGGATAGATGTCCCC..... 866

393 MetThrProProTyProGlnPheGluGlnSerGluThrGluThrValHi 409
866 ..... 866

409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyG 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
867 .....ATTCCAGATTTGCTGCTTGCATGTGAATAGGAGAAATGGAG 909

426 lnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
910 AGATGATCAAAAAAATACAAAATGATGCTGTGCTGCATTCAGTTTAAG 959

443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
960 CCAGATGATGGGACAAACACCCGAA...AGGATAGCACAAATAACAGACC 1006

459 oProGlu...AlaGlnPheLysAlaGln..... 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 TCAGACCCGATGTCAACATGCTGCAGAAATATTATACAGACCTCTTCGAA 1056

468 .....GlyArgIleTyr 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 GTGTTCCAGGCTGTAACTCTGTGTGGACCTGGACCTGGTGTCCGAGGA 1106

472 GlyLys.....IleLysG 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1107 GTGATGAGGTCAAGGCAACTGGAACATGGGACCACTGGTGGATTACAGGA 1156

476 uGluAsnPheValSerProLysGluGluValLysLeuAlaHisIleA 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1157 ATTTAATTTATT..... 1169

493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1170 ..GTGCCAACTGGGAAACTGGATTAAATAATAGAAAGAGGTGAAACC 1217

510 ValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArgAs 526
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 ATAAAAGCATAAAGCCAGCAGCTGTGTGCAAGAAATAGAACTTCAGAGAAA 1267

526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1268 TCTCCACCAATGCAGATCTCTAATGAAGTTATTACAAATTCGTGGCA 1317

542 is.....PheTyAlaCysGlnValAlaGlnArgLysIle 553
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1318 CTCACAACAGATAGACTATGCTCGCAACTCATAGAGAAAGAAAGATT 1364

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-021-608D-7
seq_documentation_block:
; Sequence 7, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1803
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; US-08-021-608D-7

```

```

alignment_scores:
  Quality: 229.50      Length: 405
  Ratio: 1.136         Gaps: 16
  Percent Similarity: 49.877   Percent Identity: 22.469

```

```

alignment_block:
US-09-685-696-176 x US-08-021-608D-7

```

```

Align seg 1/1 to: US-08-021-608D-7 from: 1 to: 1803

```

```

202 ValProThrGlnPheValGlyAlaIleGlyLysGluGlyAlaThrIl 218
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
295 GTTCCAGATGAATGGTGGATTGATTAATGGCAGAGGAGGTGAACAGAT 344
218 eArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysG 235
| :||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
345 CTCACGCATACACAGGAATCTGGATGCAAAATACAGATA...GCTCTG 391
235 luAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerProGlu 251
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
392 ACAGTGTGGCTTCCAGAAAGGTCCTGTATGTTAACTGGGAACACCTGAA 441
252 GlyThrSerAlaAlaCysLysSerIleLeuGluIleMethHisLysGluAl 268
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
442 TCTCTCAGTCAGCAAAACGGTACTGGACAGATTGTTGAAAAGGAAG 491
268 aGlnAspIleLysPheThrGlu.....GluIleProLeuL 280
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
492 ACCAGCTCTGGCTTCCATCATGCGGATGACCGGGAATGCAGTTCAAG 541
280 ysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGly 296
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
542 AAATCATGATTCCAGCTAGCAAGGAGGATAGTCATTGGAAAAGGGGA 591
297 ArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSe 313
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
592 GAACATATTAAACAGCTTCAGGAACGGGCTGAGTAAATGGTTATG.. 639
313 rProLeuGlnGluLeuThrLeuThrAsnProGluArgThr..... 326
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
640 ....ATTCAAGAC.....GGGCGCAGCAACACTGGTGCTGACA.. 673

```

```

327 .....IleThrValLysGlyAsnValGluThrCysAlaLysAlaGluGlu 341
::: :|||:||||| :|||:||||| :|||:||||| :|||:|||||
674 AACCTCTTAGGATTACAGGAGACCCATATAAGTTCAACAGCAAGGAA 723
342 GluIleMetLysLysIleArgGlu.....SerTyrGluAs 353
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
724 ATGGTGTTAGAGTTAAATTCGTGATCAAGCGGTTTCAGAGAAGTTCGGA 773
353 nAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnL 370
|::: :|||:||||| :|||:||||| :|||:||||| :|||:|||||
774 TCAGTATGGGTCAGA.....ATAGGAGA..... 798
370 euAsnAlaLeuGlyLeuPheProThrSerGlyMetProProThr 386
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
799 .....AATGAAGGATAGATGTCCCC... 819
387 SerGlyProProSerAlaMetThrProProTyrProGlnPheGluGlnSe 403
819 ..... 819
403 rGluThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaI 420
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
820 .....ATTCCAAGATTTGCTGTGGCATG 844
420 lelleGlyLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGly 436
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
845 TAATAGGAAGAAATGGAGAGATGATCAAAAAAATACAAAATGATGCTG 894
437 AlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMe 453
::: :|||:||||| :|||:||||| :|||:||||| :|||:|||||
895 GTTCGCATTCAGTTTAAGCCAGATGATGGGACACACCCGAA...AGGAT 941
453 tValIleIleThrGlyProGlu...AlaGlnPheLysAlaGln.... 467
::: :|||||:||||| :|||:||||| :|||:||||| :|||:|||||
942 AGCACAATAACAGACCTCCAGACCGATGTCACATCTCGAGAAATTA 991
467 ..... 467
992 TTACAGACCTTCTTCAAGTGTTCAGGCTGTAATCTGCTGGACCTGGA 1041
468 .....GlyArgIleTyrGlyLys..... 473
1042 CCTGGTGTTCGAGGAAGAGGTAGAGGTCAAGGCAACTGGAACATGGGACC 1091
474 .....IleLysGluGluAsnPheValSerProLysGluGluValL 487
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1092 ACCTGGTGGATTACAGGAATTTAATTTATT..... 1122
487 ysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIle 503
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1123 .....GTGCCAACTGGGAAACCTGGATTATAATA 1152
504 GlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGl 520
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1153 GGAAAGGAGGTGAACACCATAAAAGCATATAAGCCAGCAGTCTGGTCAAG 1202
520 uValValValProArgAspGlnThrProAspGluAsnAspGlnVal... 536
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1203 AATAGAATTCAGAGAATCTCCACCAATATGCAGATCCTTAATATGAAGT 1252
536 alValLysIleThrGlyHis.....PheTyrAlaCysGlnVal 548
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1253 TATTTACAATTCGTGGCAGCTCCACAACAGATAGACTATGTCGGCACTC 1302
549 AlaGlnArgLysIle 553
:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1303 ATAGAAGAAAGATT 1317

```

```

seq_name: /cqn_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-7

```

```

seq_documentation_block:
; Sequence 7, Application US/08726160

```


OM of: US-09-685-696-176 to: EST:* out_format : pfs

Date: Jan 25, 2002 10:46 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-Q/cgnt2/_uspto_spool/US09685696/runat_25012002_145550_10634/app_query.fasta_1.648
-DB=EST -OFT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCI=0.000 -LOAPEXT=0.000 -QGAPOP=4.000
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTPMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09685696 -CGNT1_1_4152 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-685-696-176

Query length: 579

Database: EST:*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1537.690000

score_list:

| Sequence | Strd | Orig | zScore | Escore | Len | Documentation |
|-------------------|------|---------|---------|----------|------|----------------------------------|
| gb_hlc:AK011689 | + | 2858.00 | 4213.85 | 1.5e-225 | 2202 | AK011689 Mus musculus 10 days |
| gb_hlc:AK013940 | + | 2192.00 | 3225.42 | 1.7e-170 | 2780 | AK013940 Mus musculus 13 days |
| gb_est2:BF080004 | + | 1155.00 | 1702.09 | 1.2e-85 | 769 | BF080004 WTL_000946 Mouse 9-day |
| gb_est1:AL127673 | + | 1065.50 | 1570.01 | 9.2e-78 | 733 | AL127673 AUI27673 NT2RP2 Homo S |
| gb_est1:BE345335 | + | 1047.50 | 1542.64 | 9.2e-77 | 730 | BE345335 601070391F1 NIH_MGC_12 |
| gb_est1:AL584627 | + | 975.50 | 1437.09 | 4.0e-71 | 707 | AL584627 AL584627 Stratagene CH |
| gb_est2:BG079106 | + | 952.00 | 1404.06 | 4.8e-69 | 587 | BG079106 H3035G11-5 NIA Mouse 1 |
| gb_est1:AA117282 | + | 936.50 | 1379.75 | 1.1e-67 | 677 | AA117282 mm25a10.r1 Beddington |
| gb_est1:AA163596 | + | 911.00 | 1342.45 | 1.3e-65 | 645 | AA163596 mm38902.r1 Beddington |
| gb_est1:AL584375 | + | 909.50 | 1340.20 | 1.7e-65 | 647 | AL584375 AL584575 Stratagene CH |
| gb_est1:BF984377 | + | 903.00 | 1331.52 | 5.3e-65 | 586 | BF984377 AV6123377 Bos taurus lu |
| gb_est2:BF080004 | + | 890.00 | 1308.22 | 1.1e-63 | 896 | BF080004 602308484F1 NIH_MGC_88 |
| gb_est1:AL34323 | + | 884.00 | 1301.09 | 2.6e-63 | 745 | AL34323 AUI34323 OVARC1 Homo s |
| gb_est2:BG082113 | + | 868.00 | 1279.90 | 4.0e-62 | 573 | BG082113 H3072F04-5 NIA Mouse 1 |
| gb_est1:AA691146 | + | 835.50 | 1232.01 | 1.8e-59 | 559 | AA691146 vt68h09.s1 Knowles Sol |
| gb_est1:AL584370 | + | 831.50 | 1225.32 | 4.4e-59 | 606 | AL584370 AL584570 Stratagene CH |
| gb_est2:BF217988 | + | 822.50 | 1210.91 | 2.8e-58 | 679 | BF217988 601862766F1 NIH_MGC_57 |
| gb_est1:AW342998 | + | 822.50 | 1210.89 | 2.8e-58 | 680 | AW342998 fi72a06.y1 Sugano Kawa |
| gb_est2:BG748346 | + | 816.50 | 1199.46 | 1.2e-57 | 889 | BG748346 602705902F1 NIH_MGC_43 |
| gb_est2:BE069090 | - | 814.50 | 1198.57 | 1.3e-57 | 715 | BE069090 H3072F04-3 NIA Mouse 1 |
| gb_est2:BE869656 | + | 803.00 | 1179.33 | 1.6e-56 | 902 | BE869656 601446879F1 NIH_MGC_65 |
| gb_est1:AA349384 | + | 791.00 | 1167.30 | 7.4e-56 | 494 | AA349384 vk86003.s1 Knowles Sol |
| gb_gss:AF046624 | + | 776.50 | 1142.22 | 1.9e-54 | 721 | AF046624 AF046624 Mus musculus |
| gb_est1:AA116751 | + | 763.00 | 1125.00 | 1.7e-53 | 539 | AA116751 mm21h01.r1 Beddington |
| gb_est1:AA587050 | + | 753.50 | 1109.60 | 1.2e-52 | 620 | AA587050 AL587050 Stratagene CH |
| gb_est1:AA427803 | + | 735.00 | 1085.50 | 2.7e-51 | 439 | AA427803 zw49c11.r1 Soares tota |
| gb_est1:AL558893 | + | 726.00 | 1089.53 | 2.1e-50 | 579 | AL558893 fb79c12.y1 Zebrafish W |
| gb_est2:BG086000 | + | 721.50 | 1063.53 | 4.5e-50 | 540 | BG086000 H3120E12-5 NIA Mouse 1 |
| gb_est1:AL586757 | + | 719.50 | 1059.32 | 7.7e-50 | 616 | AL586757 AL586757 Stratagene CH |
| gb_est1:AA134724 | + | 712.00 | 1047.80 | 3.4e-49 | 643 | AA134724 AUI34724 PLACE1 Homo s |
| gb_est1:AW343989 | - | 706.00 | 1038.08 | 1.2e-48 | 702 | AW343989 fi72a06.x1 Sugano Kawa |
| gb_est2:BG107753 | + | 704.50 | 1029.50 | 3.5e-48 | 1368 | BG107753 602277962F1 NIH_MGC_8 |
| gb_est2:BG115319 | + | 684.00 | 1003.89 | 9.4e-47 | 831 | BG115319 602316224F1 NIH_MGC_86 |
| gb_est1:AAU117945 | + | 676.00 | 991.24 | 4.4e-46 | 904 | AAU117945 AUI17945 HEMBA1 Homo s |
| gb_est1:BE622021 | + | 663.50 | 974.40 | 4.1e-45 | 759 | BE622021 601440673F1 NIH_MGC_72 |
| gb_est2:BF185290 | + | 660.00 | 970.52 | 6.8e-45 | 662 | BF185290 601843593F1 NIH_MGC_54 |
| gb_est1:AA120364 | + | 658.00 | 969.04 | 8.2e-45 | 567 | AA120364 mm17b02.r1 Beddington |
| gb_est1:AA141739 | + | 648.00 | 956.24 | 4.2e-44 | 459 | AA141739 EST291818 Normalized I |
| gb_est2:C77087 | - | 635.00 | 934.45 | 6.9e-43 | 599 | C77087 C77087 Mouse 3.5-dpc bla |
| gb_est2:BE81835 | + | 633.50 | 928.25 | 1.5e-42 | 910 | BE81835 601504962F1 NIH_MGC_71 |
| gb_est2:BG499373 | + | 632.00 | 929.49 | 1.3e-42 | 633 | BG499373 602547471F1 NIH_MGC_60 |

gb_est2:BF241103 + 631.00 924.38 2.5e-42 926 ! BF241103 601880647F1 NIH_MGC
gb_est2:BF185289 + 620.50 912.15 1.2e-41 654 ! BF185289 601843592F1 NIH_MGC
gb_est2:DS8801 + 620.00 916.11 7.3e-42 399 ! DS8801 HUM512B06B Clontech h
gb_est1:AA073514 + 617.00 910.81 1.4e-41 437 ! AA073514 mm83e03.r1 Strategic
seq_name: gb_hlc:AK011689

seq_documentation_block:

LOCUS AK011689 2202 bp mRNA 05-JUL-2001
DEFINITION Mus musculus 10 days embryo cDNA, RIKEN full-length enriched
library, clone:2610036B18, full insert sequence.
ACCESSION AK011689
VERSION AK011689.1 GI:12847972
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2610036B18.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Carninci,P. and Hayashizaki,Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

MEDLINE

PUB


```

/protein_id="BAB29071.1"
/db_xref="GI:12851514"
/translation="MNKLYIGNLNESVTPADLEKVPFAEHKHSYSGQFLVKSGYAFVDC
PDEHMAKAEFTSGVLEQGRLEIEHSVPKQKRSKTOIRNIPPQLRMEVLDLSLLA
QTGVNMCQVNTSESTAVVNTYSNREQTQRAIMKLNGHLENHALKVSYIPDEQIT
QGPENRRGGFSRGOPROGSPVAAGAPAKQOPVDIPLRLVPTQYVGAIGKEGATI
RNITKQTSKIDVHRKENAGAEKATSVHSTPEGCSSACKMILEIMHKEAKDTKTADE
VPLKILAHNNFVGRLLIGKEGRNLKQVEQDTEFKITISSLDLTLNPERTIYVKGAE
NCCRAQDEIMKKVREAYENDVAMSLSHLIPGLNAAVGLFPASSAVPPPPSSVTG
AAPTYSFQMAQPEQMVFIQAQAGAITKKGOHIKQLSRFASASIKTAPETPDSK
VMVNTYQAFQAGRIYGLKEENFPKPEVKLETHIRVPSAAGRIVIGGGK
TVNELQNLTAEEVVPDQTPDENQVIVKIIGHEFVASQMAQRKIRDLIAQVQKHQK
GOSNLAQARRK"
BASE COUNT      691 a      787 c      752 g      550 t
ORIGIN

```

```

alignment_scores:
  Quality: 2192.00      Length: 585
  Ratio: 4.152          Gaps: 8
  Percent Similarity: 90.256      Percent Identity: 74.188

```

alignment_block:

US-09-685-696-176 x AK013940

Align seg 1/1 to: AK013940 from: 1 to: 2780

```

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaProSerAs 17
|||||
279 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGTGTGACCCCGCAGA 328
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
329 CTTGGAGAAAGTATTCGGGAGACAAAGATCTCTACAGCGCGCAGTCT 378
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
379 TGGTCAATCCGGCTACGCCCTTCGTGATTGCCCGGACGACACTGGGCG 428
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
429 ATGAAGGCCATCGAAACTTTCTCGGGAAAGTAGAACTGCAAGGAAACG 478
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
479 TCTAGAGATTGAACACTCAGTCCCAAAAACAAAGGAGTGGGAAATAC 528
84 InileArgAsnIleProHisLeuGlnTrpGluValLeuAspSerIleu 100
|||||
529 AGATTCCGATATTCACCTCAGCTCGATGGGAAGTGCTAGATAGCCTG 578
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
579 CTGGCTCAGTACGCTACAGTGGGAACTGTGAGCAAGTGAACACTGAAG 628
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
629 TGAGACAGCTGTGTGTCACAGCTCCTACTACTAACCGGAGCAGCAGGC 678
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
679 AAGTATCATGAACCTAATAGGCCATCACTGGAGAACCATGCCCTGAAG 728
151 ValAlaTyrIleProAspGluMetAlaIleGlnGlnAsnProLeuGlnG 167
|||||
729 GTCTCCTACATACCTGATGAGCATAACG...CAAGGTCCTGAGAAATGG 775
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly 184
|||||
776 GCGTCGTGGA.....GGCTTTGGGCTCGGGGCCACCCCGCAAGGGT 819
184 erPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
|||||

```

```

820 CGCCCGTGGCAGCAGGGGCTCCAGCCRAAGCAGCAGCCAGTGGACATCCCT 869
198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIlelleGlyLysG 214
|||||
870 CTCGGGCTCTGGTGGCTACGAGTATAGGCGGTATCATTTGGCAAGA 919
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
|||||
920 GGGTGCACCACCTCGGAACATCACAAAACAGACAGCAGTCCAAAATAGAC 969
231 alHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
|||||
970 TGCATAGGAAGAGAAATCGGGCGCTCGGAGAAAGGCCATCAGCGTGCAT 1019
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 264
|||||
1020 TCAACCCCTGAAGGCTGCTCTCCGCGTGAAGATGATCTTTGGAGATTAT 1069
264 thIsLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
|||||
1070 GCACAAGGAGGCAAGGACACCAAAACGGCAGATGAAGTTCCCTGAAGA 1119
281 leLeuAlaHisAsnAsnAsnAsnValGlyArgLeuIleGlyLysGluArg 297
|||||
1120 TCCTGGCTCATACAACCTTCGCGGCGACTCATTTGGCAAGAGGGCGG 1169
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
|||||
1170 AACCTGAAGAGGTGGAGCAGCAGACAGACAGAGATCACCATCTCATC 1219
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
|||||
1220 GCTCCAGGACCTCAGCTCTATAACCTCGAGAGGACCATCATCTGTGAAG 1269
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||
1270 CGGCATTGAGAACTGTTCAGGCGCCGAGGAGATCATGAAGAAAGTT 1319
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
|||||
1320 CGAGAGCTTACGAGAAACGAGCTGGCGCCATGAGCTTGAGTCCACCT 1369
364 uileProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerG 381
|||||
1370 CATCCCTGGGCTTAACCTGGCTGTAGTCTCTCCAGCTTCATCCA 1419
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
|||||
1420 CGCGTGTCCCTCCT.....CCTCCAGCAGTGTCCAGGGGCTGCT 1460
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
|||||
1461 CCCTATAGTCTCTTCATGAGGCTCCGAGCAGGAGATGGTACAAGTGT 1510
411 eileProAlaLeuSerValGlyAlaIlelleGlyLysGlnGlnHisI 428
|||||
1511 CATCCCGCCCGAGGCTGTGGCGCCATCATTTGCAAGAGGGCCAGACA 1560
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
|||||
1561 TCANAACACTCTCCCGTTCGCCAGCGCTCCATCAAGATTCACACCACA 1610
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProG 461
|||||
1611 GAAACACTGACTCAAAAGTTCGAATGCTGCTCATCTACTGGACCCCGAGA 1660
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGlu 478
|||||
1661 GGCTCAGTTCAAGGCCCGAGGAAATCTATGGCAAACTAAAGAGAGA 1710
478 snPheValSerProLysGluValLysLeuGluAlaHisIleArgVal 494
|||||
1711 ATTTCTTTGGTCCCAAGGAGAGTAAGCTAGAGACCCACATACGGGTT 1760

```

495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAs 511
 1761 CCGGCTTCAGACGCGCGTGTATCGCAAGGCGGCAAAAGCGTGAA 1810
 511 nGluLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnT 528
 1811 TGAGCTGCAGAACTTTCAGCGACGCTAGTGTAGTGCACCAAGAGCAGA 1860
 528 hrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyr 544
 1861 CCCCAGTACAGAACGACCAAGTCATTTGTTAAGATCATCGGACATTTCTAT 1910
 545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnVal 561
 1911 GCCAGCCAGATGGCTCAGCGGAAGATCCGAGACATCCTGGCTCAAGTTAA 1960
 561 sGlnHisGlnGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgA 578
 1961 GCAA...CAGCACCAAGAGGA...CAGAGCAACCTGGGCCCGAGGCTCGGA 2004
 578 rgLys 579
 2005 GGAAG 2009

seq_name: gb_est2:BF608004

seq_documentation_block: 769 bp mRNA EST 01-APR-2001
 LOCUS BF608004 MY1_000946 Mouse 9-day fetus cdna library ICRPp522 Mus musculus
 DEFINITION CDNA clone ICRFP522M2029 5', mRNA sequence.
 ACCESSION BF608004
 VERSION BF608004.1 GI:13504496
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 769)
 Yahyavi, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B.G.,
 Lehrach, H. and O'Brien, J.
 TITLE Detection of a high number of novel genes in a 9-day mouse embryo
 cdna library normalised by oligonucleotide fingerprinting
 JOURNAL Unpublished (2001)
 COMMENT Contact: Hennig S
 Laboratory 123, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihnestr.63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1612
 Fax: +49 30 8413 1380
 Email: hennig@molgen.mpg.de
 EST's are made from clones being representatives of clone clusters.
 Clone clusters were calculated from oligonucleotide fingerprints.
 PCR Primers
 FORWARD: 5'-CAGCTATTCCAGAAGTAGTGA-3'
 BACKWARD: 5'-TAATAGCCTCACTATAGGG-3'
 Seq primer: 5'-ATTAGGTGACACTATAG-3'
 High quality sequence stop: 769.
 FEATURES
 Location/Qualifiers
 1..769
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="ICRFP522M2029"
 /clone_lib="Mouse 9-day fetus cdna library ICRPp522"
 /tissue_type="whole embryo"
 /dev_stage="embryonic 9-day"
 /lab_host="E.coli, XL1 blue"
 /note="Vector: pSVSPORT1; Site 1: NotI; Site 2: SalI;
 Library preparation by oligo-dT priming of RNA. Clones can
 be ordered from the Resource Center in Berlin,
 http://www.rzpd.de."
 208 a 195 c 191 g 174 t 1 others
 BASE COUNT
 ORIGIN

alignment_scores:
 Quality: 1155.00 Length: 242
 Ratio: 4.915 Gaps: 1
 Percent Similarity: 97.107 Percent Identity: 94.628
 alignment_block:
 US-09-685-696-176 x BF608004 ..
 Align seg 1/1 to: BF608004 from: 1 to: 769
 316 GlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAs 332
 45 CAGGAATCAGCGTGTACAAATCCGGAACGACCATACAGTGAAGGCAG 94
 332 nValGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgG 349
 95 TGTGTGAGACGTGTGCCAAGCGGAGGAGGAATAATGAAGAAGATCAGG 144
 349 luSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIle 365
 145 AGTCTTATGAAATGATATGTCTTCCATGAATCTTCAAGCACATTTAATC 194
 366 ProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMe 382
 195 CCTGGATTAAATCTGAATGCCTTGGGTCTTCCACCCACGTCAGGGAT 244
 382 tProProProThrSerGlyProProSerAlaMetThrProProTyrProG 399
 245 GCACCTCCACCTCAGGCGCCCTTCAACCTGACTCTCCCTACCCAC 294
 399 inPheGluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeu 415
 295 AATTTCAGCAATCAGACGACGACTGTGCATCTGTTATTCGCGCCTG 344
 416 SerValGlyAlaIleLeuGlyLysGlnGlnHisIleLysGlnLeuSe 432
 345 TCCGTTGGCGCCATCATTTGCCAAGCAGGCGCAACACATCAACAGCTTC 394
 432 rArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspA 449
 395 TCCTTTGGGAGCTTCGATTAAAGATCGCTCCAGCGAAGCAGCAGATG 444
 449 laLysValArgMetValIleIleThrGlyProProGluAlaGlnPheLys 465
 445 CTAAAGTCGGATGGGATGATATCCTGGACCCAGCAGAGGCTCAGTTCAAG 494
 466 AlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerPr 482
 495 GCTCAGGGAAGAAATTTATGCAAAATTTAAAGAAACACACTTTGTTAGTCC 544
 482 oLysGluGluValLysLeuGluAlaHisIleArgValProSerPheAlaA 499
 545 TAAAGAAGAGGTGAACCTTGAAGCTCAGATCAGAGTCCCTCTTCTGCTG 594
 499 laGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsn 515
 595 CTGGCAGATTTATGGGAAGAGGCAAAACTGTGAATGAGTCCAGAGT 644
 516 LeuSerSerAlaGluValValProArgAspGlnThrProAspGluAs 532
 645 TTATCAGCTGTTGAGNTGTCGTCCTCCCGTGACACACCTGATGAGAA 694
 532 nAspGlnValValLysIleThrGlyHisPheTyrAla.CysGlnVal 548
 695 TGATCAAGTAGTGTCAACATAACTGGCCACTTCTATCTATCTTGGCCGGGT 744
 549 AlaGlnArgLysIleGlnGluIle 556
 745 GCCCAGAGGAAATTCAGGAAT 768
 seq_name: gb_est1:AU127673

```
seq_documentation_block:
  LOCUS      AUI27673      733 bp      mRNA      EST      24-OCT-2000
  DEFINITION AUI27673 NT2RP2 Homo sapiens cDNA clone NT2RP2001818 5', mRNA
  sequence.
  ACCESSION  AUI27673
  VERSION    AUI27673.1 GI:10988027
  KEYWORDS   EST.
  SOURCE     human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE  1 (bases 1 to 733)
  AUTHORS    Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
              Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
              Isogai,T.
  TITLE      HRI human cDNA project
  JOURNAL    Unpublished (2000)
  COMMENT    Contact: Takao Isogai
              Genomics Laboratory
              Helix Research Institute
              1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              Tel: 81-438-52-3951
              Fax: 81-438-52-3952
              Email: genomics@hri.co.jp
              HRI human cDNA project; 5' & 3'-end one pass sequencing; Helix
              Research Institute; cDNA library construction; Department of
              Virology, Institute of Medical Science, University of Tokyo, and
              Helix Research Institute.
  FEATURES   Location/Qualifiers
              source          1..733
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="NT2RP2001818"
                        /clone_lib="NT2RP2"
                        /cell_type="teratocarcinoma"
                        /cell_line="NT2"
                        /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                        cells after 2-weeks retinoic acid (RA) induction"
  BASE COUNT      212 a 181 c 148 g 180 t 12 others
  ORIGIN

alignment_scores:
  Quality: 1065.50      Length: 230
  Ratio: 4.910          Gaps: 2
  Percent Similarity: 94.348      Percent Identity: 93.913

alignment_block:
  US-09-685-696-176 x AUI27673 ..
  Align seg 1/1 to: AUI27673 from: 1 to: 733
  194 CysAspLeuProLeuArgLeuLeuValProThrGlnPheValGlyAlaI 210
  |||||
  5 TGTGATTGGCTCTGCGCCTGCTGGTTCACCAACCAATTTGTTGGAGCAT 54
  |||||
  210 eileGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlns 227
  |||||
  55 CATAGGAAAGAAGAGTCCACCATTCGGACATCCCAACACAGACCAGT 104
  |||||
  227 erTysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSer 243
  |||||
  105 CTAATAATCGATGTCACCGTAAGCAAAATGCGGGGCTGCTGAGAACTCG 154
  |||||
  244 IleThrIleLeuSerThrProGluGlyThrSerAlaAlaLysLysSerI 260
  |||||
  155 ATTAATACTCTCTACTCTCTGAGGACACCTCTGCGGCTTTGTAAGTCTAT 204
  |||||
  260 eLeuGluIleMethIstLysGluAlaGlnAspIleLysPheThrGluGlu 277
  |||||
  205 TCTGGAGATTATGCTAAGGAGCTCAAGATATAAAATTCACANAAGAGA 254
  |||||
```

```
277 leProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGly 293
|||||
255 TCCCTTTGAAGATTTAGCTCATAATAACTTTGTTGGACGCTTATTGGT 304
|||||
294 LysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLysI 310
|||||
305 AAAGAAGGAAGAAATCTTAAAAAATTGAGCAAGACACAGACNCTAAAT 354
|||||
310 eThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluArgThrI 327
|||||
355 CACATATCTCCATTGCAGGAATTGACGCTGTATTAATCCAGAACGCACTA 404
|||||
327 leThrValLysGlyAsnValGluThrCysAlaLysAlaGluGluIle 343
|||||
405 TTACAGTTAAAGGCAATGTTGAGACATCTGCCAAAGCTGAGGAGGATC 454
|||||
344 MetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMetAsnI 360
|||||
455 ATGAAGAANATCAGGAGTCTTATGAAAATGATATTGCTTCTGAATCT 504
|||||
360 uGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyLeuPheP 377
|||||
505 TCNAGCACATNTAATTCCTGNATTAAATCTGAAGCCCTTGGGCTGTGTC 554
|||||
377 roProThrSerGlyMetProProThrProThrSerGlyProProSerAlaMet 393
|||||
555 CACCACCTTCAGGATGCACCTCCACCTCCAGGCCCTTCAGCCATG 604
|||||
394 ThrProProTyrProGlnPheGluGlnSerGluThrGluThr.ValHISG 410
|||||
605 ACTCCTCCCTCCGCCGAGTTTGAGCAATCCGAAACGGAGACTGGTTCNTC 654
|||||
410 lnPheIleProAlaLeu...SerValGlyAlaIleIle 421
|||||
655 TGTTTTATCCCCACCTCTATCCAGTCGGTGCCATCATC 692

seq_name: gb_est1:BE545535
seq_documentation_block:
  LOCUS      BE545535      790 bp      mRNA      EST      09-AUG-2000
  DEFINITION 601070391F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456407 5',
  mRNA sequence.
  ACCESSION  BE545535
  VERSION    BE545535.1 GI:9774180
  KEYWORDS   EST.
  SOURCE     human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE  1 (bases 1 to 790)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
  COMMENT    Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              plate: LLAM8444 row: f column: 24
              High quality sequence stop: 667.
              Location/Qualifiers
              source          1..790
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:3456407"
                        /clone_lib="NIH_MGC_12"
                        /tissue_type="cervical carcinoma cell line"
                        /lab_host="DH10B"
                        /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
```

Site_2: Sali; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 221 a 193 c 209 g 167 t
ORIGIN

alignment_scores:
Quality: 1047.50 Length: 259
Ratio: 4.476 Gaps: 4
Percent Similarity: 90.347 Percent Identity: 85.714
alignment_block:
US-09-685-696-176 x BE545535 ..
Align seg 1/1 to: BE545535 from: 1 to: 790

29 ValSerGlyProPheLeuValLysThrGlyTyrAlaPheValAspCysPr 45
|||||
3 GTGTCGGGACCTCTCTGTTGAAGACTGGCTAGCGTTCGTGACTGCC 52
|||||
45 oAspGluSerTrpAlaLeuLysAlaIleGluAlaLeuSerGlyLysIleG 62
|||||
53 GGACGAGAGCTGGCCCTCAAGCCCTCAGGCGCTTCAGGTAAATAG 102
|||||
62 lueLueHisGlyLysProIleGluValGluHisSerValProLysArgGln 78
|||||
103 AACTGCACGGGAACCATAGAGTTGAGCACTCGTCCCAAAAGGCAA 152
|||||
79 ArgIleArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpG 95
|||||
153 AGGATTCGGAACCTCAGATACCAATATCCCGCCTCATTTACAGTGGGA 202
|||||
95 uValLeuAspSerLeuLeuValGlnTyrGlyValValGluSerCysGluG 112
|||||
203 GGTCTGGGATGTTTACTAGTCAGATGGAGTGGTGGAGAGCTGTGAGC 252
|||||
112 InValAsnThrAspSerGluThrAlaValValAsnValThrTyrSerSer 128
|||||
253 AACTGACACTGACTCGGAACCTGCAGTTGTAATGTAACTATTTCAGT 302
|||||
129 LysAspGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuG 145
|||||
303 AAGGACCAAGCTAGACAAGCACTAGACAACACTGAATGGATTTTCAGTTAG 352
|||||
145 uAsnPheThrLeuLysValAlaTyrIleProAspGluMetAlaAlaGlnG 162
|||||
353 GAATTTTCACCTTGAAAGTAGCCCTATATCCCTGATGAATGGACGCCAGC 402
|||||
162 InAsnProLeuGlnGlnProArgGlyArgGlyLeuGlyLeuGlnArgGly 178
|||||
403 AAAACCCCTTCAGACAGCCCGGAGTGGCGGGGCTTGGGCAGAGGGGC 452
|||||
179 SerSerArgGlnGlySerProGlySerValSerLysGlnLysProCysAs 195
|||||
453 TCCTCAAGGCAGGGTCT. CCAGATCCGTATCCCAAGCAGAAACCATGTGA 501
|||||
195 pLeuProLeuArgLeuLeuValProThrGlnPheValGlyAlaIleIle. 211
|||||
502 TTTGCTCTCGCGCTGCTGTTGCCACCCCAATTTGTTGGAGCCATCATAG 551
|||||
212 GlyLysGluGlyAlaThrIleArgAsnIle. ThrLysGlnThrGlnSerL 228
|||||
552 GGAAGAAGAGTGGCCACCATTCCTGGACATCTACCAACAGACAGCCAGTCA 601
|||||
228 ysIleAspVal. HisArgLysGluAsnAlaGlyAlaAlaGluLysSerI 244
|||||
602 AAATCATGTCTCCACCCGTGAAGAAATCGCGGCTGCTG... AGAAGTCG 648
|||||
244 eThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleL 261
|||
649 ACTACTATCTCTCTATCTGAAGGAGCCTCTGGGATTTGTAAGACTATCCT 698
|||

261 euGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluGluIle 277
|||||
699 GGGAATATTGCTAAGGAGGCGCAGATTTTAAATCCAGAGAGAATC 748
|||||
278 ProLeuLysIleLeuAlaHis 284
|||||
749 CCTGTAATAATTTATGCCCAT 769
seq_name: gb_est1:AL584627
seq_documentation_block:
LOCUS AL584627 707 bp mRNA EST 28-FEB-2001
DEFINITION AL584627 Stratagene Chick Embryo Lambda cDNA Library (* 937405)
Gallus gallus cDNA clone ROS012C05, mRNA sequence.
ACCESSION AL584627
VERSION AL584627.1 GI:13163358
KEYWORDS EST.
SOURCE Chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 707)
Murray,F.
Stratagene Chick Embryo Lambda cDNA Library
TITLE Unpublished (2001)
JOURNAL Contact: Frazer Murray
COMMENT Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
Seq primer: T3.
FEATURES
Location/Qualifiers
1..707
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="ROS012C05"
/clone_lib="Stratagene Chick Embryo Lambda cDNA Library (*
937405)"
/tissue_type="Embryo"
/dev_stage="5 days old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Vector: pBLUESCRIPT SK; Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dt. Uni-ZAP XR
vector. Average insert size: 1.5kb.; 5' adaptor sequence:
5' GAATTCGGCAGCAG 3'; 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 218 a 150 c 159 g 180 t
ORIGIN
alignment_scores:
Quality: 975.50 Length: 231
Ratio: 4.495 Gaps: 4
Percent Similarity: 93.939 Percent Identity: 85.714
alignment_block:
US-09-685-696-176 x AL584627 ..
Align seg 1/1 to: AL584627 from: 1 to: 707
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
|||||
17 AACCTGAAGAAATTTGACGAGGACACACTATAAAATCAGATATCTCC 66
|||||
314 oLeuGlnGluThrLeuTyrAsnProGluArgThrIleThrValLysG 331
|||||
67 ATTGCAGGACTTGACACTATATAATCCGGAAGGACCATACAGTTAAG 116
|||||
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||


```
117 CAGTATTGAACCTTGTGCCAAGCCGAGGAAGAAATATATGAAGAAATC 166
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
167 AGGGAATCTATGAAATGATATGCTGCTATGAATCTTCAAGCACATTT 216
364 uileProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProPro...Thrs 380
217 AATTCTCGATTAATCTGAATGCTTGGTCTGTGTCACACCTCTCTCT 266
380 erGlyMetProPro...ThrsGlyProProSerAlaMetThrPro 395
267 CAGAAATACCACCTCTCCAGCTGAGTGTGCTCTGCTGCTGCTGCT 316
396 ProTyrProGlnPhe...GluGlnSerGluThrGluThrValHisGlnPh 411
317 TCATATCGCCCATTTGGSCAACACGCCGAGTCTGAGACTGTTTCATCT 366
411 eileProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyHisI 428
367 CATCCAGCCTTGGCAGTTGGAGCTATTATTGGCAAGCAGGACAGACA 416
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
417 TCAACAACATTTCTGCTTTGCAGCGGATCTATTAGATTGCCCTGCT 466
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProG 461
467 GAAGCACCAGATGCCAAGCTCAGAAATGTTATCATCACTGCACTCCAGA 516
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluAla 478
517 AGCTCAGTTTAAGGCTCAAGGAGAACTATGTAAGAACTTAAAGAAAGA 566
478 snPheValSerProLysGluValLysLeuGluAlaHisIleArgVal 494
567 ATTCTTTGGACCTTAAGAAAGAGTGAACCTTGAGGCTCATATAAAGTG 616
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThrsValas 511
617 CCATCTTATGCTGCTGGTAGAGTTATTGGTAAAGGAGGCAAAACAGTAA 666
511 nGluLeuGlnAsnLeuSerSerAlaGluValValPro 524
667 TGAGCTTCAAAACTTGACAAGTGGCAGAAGTTGTAGTCCCT 707
```

seq_name: gb_est2:BG079106

```
seq_documentation_block:
LOCUS      BG079106      587 bp      mRNA      EST      26-JAN-2001
DEFINITION H3035G11-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION  BG079106
VERSION    BG079106
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 587)
AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL   Unpublished (2001)
COMMENT   Other_ESTs: H3035G11-3
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3035 row: G column: 11
```

Seq primer: -21M13 Reverse
High quality sequence stop: 587
POLYA=No.

FEATURES

Location/Qualifiers

source

1..587
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3035G11"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"

/notes="Vector: pSPORT1; Site.1: Sall; Site.2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 156 a 152 c 135 g 143 t 1 others
ORIGIN

alignment_scores:
Quality: 952.00 Length: 195
Ratio: 5.011 Gaps: 0
Percent Similarity: 97.436 Percent Identity: 94.872
alignment_block:
US-09-685-696-176 x BG079106 ..
Align seg 1/1 to: BG079106 from: 1 to: 587

```
353 AsnAspIleAlaSerMetAsnLeuGlnAlaHisIleProGlyLeuAs 369
1 AATGATATTGCTTCCATGAATCTTCAAGCACATTTAATCCCTGGATTAA 50
369 nLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProPro 386
51 TCTCAATGCCCTTGGTCTGTTCACCACCCACGTCAGGGATGCCACCC 100
386 hrSerGlyProProSerAlaMetThrProTyrProGlnPheGluGln 402
101 CCTCAGGCCCCCTTCAACCTGACTCTCTCCCTACCCACAATTTGAGCA 150
403 SerGluThrGluThrValHisGlnPheIleProAlaLeuSerValGly 419
151 TCAGAGACGGAGACTGTGCATCTGTTATTCCCGCCCTGTCCGTGGCG 200
419 aIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAla 436
201 CATCTTGGCAAGCGGCCCAACACATCAACACAGCTTCTCGCTTGGCG 250
436 lyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysVal 452
251 GAGCTTCGATTAAAGATCGCTCCAGCCGGAAGCAGACAGATGCTAAAG 300
453 MetValIleIleThrGlyProProGluAlaGlnPheLysAlaGlnGly 469
```


301 ATGGTGATTATCACTGGACCACAGAGGCTCAGTTCAAGCTCAGGGAAG 350
469 gletyrGlyLysIleGluGluAsnPheValSerProLysGluGluVal 486
|||||
351 AATTTATGAAAAATTAAGAGAAATTTGTTAGTCTCTAAAGAGAGG 400
486 alIysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgVal 502
|||||
401 TGAAGCTTGAAGCTACATCAGAGTCCGCTCTTCTGCTGCGATAGTT 450
503 IleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAl 519
|||||
451 ATTGCAAGAGGAGGCAAAACGCTGATGAGCTCCGGAGTTTATCAAGTGC 500
519 aGluValValProArgAspGlnThrProAspGluAsnAspGlnValVal 536
|||||
501 TGAAGTTGTCACCCCGTCACAGACACCTGATGAGATGATCAAGTTG 550
536 alValLysIleThrGlyHisPheTyrAlaCysGln 547
|||||
551 TTGTACACATACTGGCCACTTCTATGCTTGCCAG 585

seq_name: gb_estl:AA117282

seq_documentation_block: 677 bp mRNA EST 15-NOV-1996
LOCUS AA117282
DEFINITION mm25a10.r1 Beddington mouse embryonic region Mus musculus cDNA
clone IMAGE:538938 5' similar to gb:L35599 Mus musculus Y-box
binding protein mRNA, 3' end (MOUSE);, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AA117282
AA117282.1 GI:1672303

ORGANISM

Mus musculus

house mouse

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 677)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

THE WASHU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:325874

Seq primer: -40ml3 ET

High quality sequence stop: 411.

Location/Qualifiers

1..677

FEATURES

source

/organism="Mus musculus"

/strain="C57BL6 x DBA"

/db_xref="taxon:10090"

/clone="IMAGE:538938"

/clone_lib="Beddington mouse embryonic region"

/sex="pooled"

/tissue_type="embryo"

/dev_stage="7.5dpc"

/lab_host="DHI2S"

/note="Organ: whole embryo; Vector: pCMV-SPORT; Site:1;

SalI; Site:2: NotI; Cloned unidirectionally. Primer:

Oligo dt. Gastrulating embryos were collected at 7.5dpc

from C57BL6 x DBA matings, excluding embryos that had

developed head folds and all extraembryonic tissues.

Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

BASE COUNT 196 a 171 c 152 g 158 t
ORIGIN
alignment_scores: Referenced in Development 121, 2479-2489 (1995)*
Quality: 936.50 Length: 219
Ratio: 4.591 Gaps: 2
Percent Similarity: 93.151 Percent Identity: 86.758

alignment_block:

US-09-685-696-176 x AA117282

Align seg 1/1 to: AA117282 from: 1 to: 677

215 GlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspVa 231

|||||

1 GGTGCTACCATTCGCAACATCACAAAGCAGACCCAGCTCTAAATCGATGT 50

|||||

231 lHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeus 248

|||||

51 CCATCGTAAGGAGATACAGGGCGCGGAGAAAGTCATTACTATCTCT 100

|||||

248 erThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMet 264

|||||

101 CTACCCCTGAAGGCACCTCTGCAGCTTGTAAAGTCTATTCTGGAGATTATG 150

|||||

265 HisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIl 281

|||||

151 CATAGAAGAGCTCAAGATATAAAATTCACAGAGAGATTCCTCTTGAAGAT 200

|||||

281 eLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgA 298

|||||

201 CTTAGCTCAACAATACTTTGTAGCCGCTCTCATTGGTAAAGAAGGAAGAA 250

|||||

298 snLeuLysLysIleGluGlnAspThrAspThrLysLysIleThrIleSerPro 314

|||||

251 ACCTTAAAAAATTCGACGACGACGACACTAAATATCAACAATATCTCCA 300

|||||

315 LeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGl 331

|||||

301 TTCGAGAACTGACGCTGTACAAATCCGAAACGCCACCATTTACAGTGAAGG 350

|||||

331 yAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIleA 348

|||||

351 CACTGTTGACACGCTGCCAAGCGGAGGAGGAAATAATGAAGAAGATCA 400

|||||

348 rGluSerTyrGluAsnAspIleAlaSerMetAsn.LeuGlnAlaHisLe 364

|||||

401 GGGAGTCTTATGAAATGATATTGCTTCCATGAATCCTTCAAGCACATTT 450

|||||

364 uIleProGlyLeuAsnLeuAlaLeuGlyLeuPheProThrSerG 381

|||||

451 AATCCCTGGATTAAATCTGAATGGCTGGTGGTCCGCCACCGTCAGG 500

|||||

381 lyMetProProThrSerSerGlyProProSerAlaMetThrProProTyr 397

|||||

501 GATGCCACTTCCACCTCAGGGGCCCTTCAACCCCTGACTCTCTCCCTAC 550

|||||

398 ProGlnPheGluGlnSerGluThrValHisGlnPheIleProAl 414

|||||

551 CCACAATTTGGCAATCAGACGCGGACGTGTCATCTGTTTATTTCCTCCG 600

|||||

414 aLeuSerValGlyAlaIleIleGlyLysGlnGlnHisIleLysGlnL 431

|||||

601 CCTGTGTTGGCGCAATCATTTGGCA...GCCGGGCCACACATTAACAGCT 647

|||||

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

431 euSer 432

648 TTTCG 652

seq_name: gb_estl:AA163596

seq_documentation_block: 645 bp mRNA EST 17-DEC-1996
LOCUS AAL163596
DEFINITION mm38g02.r1 Beddingington mouse embryonic region Mus musculus cDNA
clone IMAGE:540242 5' similar to gb:L35599 Mus musculus Y-box
binding protein mRNA, 3' end (MOUSE);, mRNA sequence.
ACCESSION AAL163596
VERSION AAL163596.1 GI:1739542
KEYWORDS EST.
SOURCE house musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 645)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMNI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:327178
Seq primer: -40m13 ET
High quality sequence stop: 465.
Location/Qualifiers
1. .645
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:540242"
/clone_lib="Beddingington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
Sali; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Gastrulating embryos were collected at 7.5dpc
from C57BL6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)."
BASE COUNT 178 a 159 c 150 g 157 t 1 others
ORIGIN

alignment_scores:
Quality: 911.00 Length: 218
Ratio: 4.401 Gaps: 3
Percent Similarity: 94.954 Percent Identity: 92.202

alignment_block:
US-09-685-696-176 x AAL163596 ..

Align seg 1/1 to: AAL163596 from: 1 to: 645

349 GluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeu11 365
|||||
3 GAGTCTTATGAAATGATATGCTTCATGAAATCTTCAGCACATTAAT 52
|||||
365 eProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyM 382
|||||
53 CCCTGGATTAATCTGAATGCTTGGGTCTGTCCACCCAGCGTCAGGA 102
|||||
382 etProProProThrSerGlyProProSerAlaMetThrProProTyrPro 398
|||||

|||||
103 TGCCACCTCCACCTCAGGGCCCCCTTCAACCCCTGACTCCCTCCCTACCCA 152
|||||
399 GlnPheGluGlnSerGluThrGluThrValHisGlnPheIleProAlaLe 415
|||||
153 CAATTTGAGCAATCAGACGAGACTGTCATCTGTTTATTCCGCGCCT 202
|||||
415 uSerValGlyAlaIleIleGlyLysGlnGlnHisIleLysGlnLeuS 432
|||||
203 GTCCGTTGGCGCCATCATTTGCAAGCAGGGCCCAACACATCAACAGCTTT 252
|||||
432 eArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAsp 448
|||||
253 CTCGCTTTGC. GGAGCTTCGATTAAAGATTCGCTCCAGCCGAAGACCAGAT 301
|||||
449 AlalysValArgMetValIleIleThrGlyProGluAlaGlnPheLy 465
|||||
302 CTTAAAGTCGGATGGTGTATTCACCTGGACACCCAGAGGCTCAGTTCAA 351
|||||
465 sAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerP 482
|||||
352 GGCTCAGGGAAGAAATTTATGAAAAAATTAAGAGAGAAACCTTTGTAGTC 401
|||||
482 rOlysGluGluValLysLeuGluAlaHisIleArgValProSerPheAla 498
|||||
402 CTTAAAGAGAGGTGAACCTTGAAGCTCACATCAGAGTCCCGCTCTTGTCT 451
|||||
499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAs 515
|||||
452 GCTGGCAGAGTATTGGGAAGAGGCAAAACGGTGAATGAGTCCAGAG 501
|||||
515 nLeuSerSerAlaGluValValProArgAspGlnThrProAspGluA 532
|||||
502 TTTATCAAGTGTCAAGTTGTCGT. CCCCGTGACCAGACACCTGATGAGA 550
|||||
532 snAspGlnValValLysLysIleThrGlyHisPheTyrAlaCysGlnVa 548
|||||
551 ATGATCAAGTAGTGTCAATAAAC...TGGCACTTCTATGCTTGCTA.GT 596
|||||
548 lAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnG 565
|||||
597 TGCCACGAGANATTTCAGA...ATTCGACTCAGTAAGCAGCAGCAGCAGC 643
|||||
565 ln 565
||
644 AG 645

seq_name: gb_estl:AL584575

seq_documentation_block: 647 bp mRNA EST 28-FEB-2001
LOCUS AL584575
DEFINITION AL584575 Stratagene Chick Embryo Lambda cDNA Library (* 937405)
Gallus gallus cDNA clone ROS011B11, mRNA sequence.
ACCESSION AL584575
VERSION AL584575.1 GI:13163306
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Murray, F.
1 (bases 1 to 647)
Stratagene Chick Embryo Lambda cDNA Library
Unpublished (2001)
Contact: Frazer Murray
Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
Seq primer: T3.

```

460 rogluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGlu 476
|||||:::|||||
565 CAGAAAGCTCAGCTTTAAGGCTCAAGGGAGAAATCTATGGAAAACCTTAAGAA 614
|||||:::|||||

477 GluAsnPheValSerProLysGluGluValLys 487
GAAAAAAAA:::|||||
615 GAAATTTCTTTGGACCTTAAGAGAAGTGAAA 647
|||||:::|||||

seq_name: gb_est1:AV612377

seq_documentation_block:
LOCUS AV612377 586 bp mRNA EST 30-AUG-2000
DEFINITION AV612377 Bos taurus lung fetus Bos taurus CDNA clone EILU045B11 5',
mRNA sequence.
ACCESSION AV612377
VERSION AV612377.1 GI:9748047
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 586)
AUTHORS Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Jitohzono,A. and
Suzuki,H.
TITLE bovine CDNA sequencing
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source
1..586
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="EILU045B11"
/clone_lib="Bos taurus lung fetus"
/tissue_type="lung"
/dev_stage="fetus"
/lab_host="DH10B"
/notes="vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 161 a 146 c 150 g 128 t 1 others
ORIGIN

alignment_scores:
Quality: 903.00 Length: 193
Ratio: 4.778 Gaps: 0
Percent Similarity: 97.927 Percent Identity: 92.746

alignment_block:
US-09-685-696-176 x AV612377 ..
Align seg 1/1 to: AV612377 from: 1 to: 586

114 AsnThrAspSerGluThrAlaValValAsnValThrTyrSerSerLysAs 130
:::|||||
2 CACGGCGTCGGCGGAACCTGCTGTTGCAACGTAACCTATTCCAATAAGGA 51
|||||

130 pGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsn 147
|||||
52 GCAAGCCACACAAGCTTTAGACAAGCTGAATGGTTCCAGCTGGAGAACT 101
|||||

147 heThrLeuLysValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsn 163
|||||
102 TCACACTGAGAGTGGCTACATCCCGGATGAGATGGCGGCCCGCCAGCAGAGC 151
|||||

```


/note="Vector: pSPORT1; Site_1: Sali; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 167 a 132 c 150 g 124 t
ORIGIN

alignment_scores:
Quality: 868.00 Length: 178
Ratio: 4.932 Gaps: 0
Percent Similarity: 98.876 Percent Identity: 97.191

alignment_block:

US-09-685-696-176 x BG082113

Align seg 1/1 to: BG082113 from: 1 to: 573

402 GlnSerGluThrGluValHisGlnPheIleProAlaLeuSerValG1 418
2 CAATCAGACAGGAGACGTGCATCTGTTTATCCCGCCCTGTCGTTGG 51
418 yAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheA 435
52 CGCCATCATTTGGCAGGAGGCGCAACACATCAACAGCTTCTCGCTTG 101
435 laGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysVal 451
102 CGGAGCTTCGATTAAAGATCGCTCCAGCTCTTTCCACGAGTCGTAAG 151
452 ArgMetValIleIleThrGlyProGluAlaGlnPheLysAlaGlnG1 468
152 CGGATGGTGATTATCGCTGGACACACAGAGGCTCAGTTCAAGGCTCAGG 201
468 yArgIleTyrrGlyLysIleLysGluGluAsnPheValSerProLysGluG 485
202 AAGAATTTATGGAATAATTAAGAAGAAACTTTGTTAGTCTCTAAAGAAG 251
485 luValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArg 501
252 AGGTGAACCTTGAAGCTCACATCAGAGTCGCTCTTTGCTGGCAGA 301
502 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSe 518
302 GTTATTGGGAAGGAGGCAAAACCGTGAATGAGCTCCAGAGTTTATCAAG 351
518 rAlaGluValValProArgAspGlnThrProAspGluAsnAspGlnV 535
352 TGCTGAAGTTGCTGCTCCCGCTGACACACACCTGATGAGAATGATCAAG 401
535 alValValLysIleThrClyHisPheTyAlaCysGlnValAlaGlnArg 551
402 TAGTTGTCAAAATTAACGTGGCCACTTCATGCTTGCACAGGTTGCCAGAG 451
552 LysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLysAl 568
452 AAATTCAGGAATTCCTGACTCAGGTAAAGCAGGACCAGCAGCAGAGAAGC 501

568 aLeuGlnSerGlyProGlnSerArgArgLys 579
502 TCTGCAGAGTCAGCACCTCAGTCAAGCGGAAG 535

seq_name: gb_estl:AA691146

seq_documentation_block:

LOCUS AA691146 559 bp mRNA EST 16-DEC-1997
DEFINITION vr68h09.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
IMAGE:1125857 5' similar to TR:000425 O00425 PUTATIVE RNA BINDING
PROTEIN KOC ; mRNA sequence.

ACCESSION AA691146

VERSION AA691146.1 GI:2692082

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lannon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:615193

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 440.

Location/Qualifiers

1..559

/organism="Mus musculus"

/strain="B6D2 F1/J"

/db_xref="taxon:10090"

/clone="IMAGE:1125857"

/clone_lib="Knowles Solter mouse 2 cell"

/tissue_type="embryo"

/dev_stage="2-cell"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pBluescribe (modified);

Site_1: MluI; Site_2: Sali; Cloned unidirectionally from

mRNA prepared from 13,500 2-cell stage embryos. Primer:

Sali(dT): 5'-CGGTCGACGCTGACGCTTTT-3',

CDNAs

were cloned into the MluI/Sali sites of a modified

pBluescribe vector using commercial linkers (NEB).

Average insert size: 1.2 kb."

BASE COUNT 154 a 139 c 155 g 111 t

ORIGIN

alignment_scores:

Quality: 835.50 Length: 188

Ratio: 4.720 Gaps: 2

Percent Similarity: 94.149 Percent Identity: 89.894

alignment_block:

US-09-685-696-176 x AA691146

Align seg 1/1 to: AA691146 from: 1 to: 559

44 CysProAspGluSerTrpAla.LeuLysAlaIleGluAlaLeuSerGlyL 60

1 TGGCCGACGACGAGCGT...GGCATCAAGGCCATCGAGCGCTTTCAGTA 47

60 ysIleGluLeuHisGlyLysProIleGluValGluHisSerValProLys 76
||||:|||||
48 AAATGGAACATACATGGGAAACCGATGGAAGTTGAGCACTCGGTCCCTAAA 97
77 ArgGlnArgIleArgLysLeuGlnIleArgAsnIleProProHisLeuGln 93
|||||
98 CGGCAGAGGATTCTGTAACCTTCAGATACGAAATATCCCGCCCCACTTACA 147
93 nTrpGluValLeuAspSerLeuLeuValGlnTyrGlyValValGluSerC 110
|||||
148 ATGGGAGGTGGTGATAGTTACTAGTCAGTATGAGTGGTGAGAGCT 197
110 ysGluGlnValAsnThrAspSerGluThrAlaValValAsnValThrTyr 126
|||||
198 CTGAGCAAGTGAACACGGATTCTGGAAACGGCAGTTGTAAATGTAACCTAT 247
127 SerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGln 143
|||||
248 TCCAGTAGGACCAAGCTAGACAGCCCTAGACAACTGAATGGATTCCA 297
143 nLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGluMetAlaA 160
|||||
298 GTTAGAGAACTTCACCTGAAGTTGCCTACATCCACAGATGAACCTGCTG 347
160 IaGlnGluAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlyGln 176
|||||
348 CCCAGCAAAATCCCTCGGCACACAGCTCCGGGGCGCCGTGGGCCAGGGCAG 397
177 ArgGlySerSerArgGlnGlySerProGlySerValSerLysGlnLysPr 193
|||||
398 CGGGGCTCATCCAGCAGCGGCTCTCCAGGATCGGTGTCCACAGCAAAACC 447
193 oCysAspLeuProLeuArgLeuLeuValProThrGlnPheValGlyAlaI 210
|||||
448 CTGTCACTGGCAGCTGGGCGCTGCTGTTCCACCAGTTTGTTCGAGCA 497
210 leIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226
|||||
498 TTATAGGAAAGAGAGTGTCTACCATTCGCAACATCACAAAGCAGACCCAG 547
227 SerLysIleAsp 230
|||||
548 TCTAGAATCGAT 559

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2002, 14:54:23 ; Search time 5806.2 Seconds
(without alignments)
11879.492 Million cell updates/sec

Title: US-09-685-696-175
Perfect score: 4181
Sequence: 1 ggtgatgcgttggtgtgt.....aaaaaaaaaaaaaaaaaaaaa 4181

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_on.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vi.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htgo_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|----------|-------|--------|----|-----------|---------------------|
| 1 | 4165 | 99.6 | 4181 | 9 | HSU971188 | U97188 Homo sapien |
| 2 | 4139 | 99.0 | 4155 | 9 | HSU76705 | U76705 Human putat |
| 3 | 3994.2 | 95.5 | 4171 | 9 | AF117108 | AF117108 Homo sapi |
| 4 | 3200 | 76.5 | 129432 | 9 | HS497321 | AL023775 Human DNA |
| c | 5 2144.6 | 51.3 | 169739 | 9 | AC005082 | AC005082 Homo sapi |
| 6 | 1511.2 | 36.1 | 1740 | 10 | AB046173 | AB046173 Mus muscu |
| c | 7 1324 | 31.7 | 181369 | 2 | AC092447 | AC092447 Homo sapi |
| c | 8 1324 | 31.7 | 188273 | 2 | AC024199 | AC024199 Homo sapi |
| 9 | 1117 | 26.7 | 3147 | 5 | AF042353 | AF042353 Xenopus l |
| 10 | 1116.4 | 26.7 | 2697 | 5 | AF064633 | AF064633 Xenopus l |
| 11 | 1084.4 | 25.9 | 161404 | 9 | AF002376 | AF002376 Homo sapi |
| 12 | 1084.2 | 25.9 | 2405 | 5 | AF064634 | AF064634 Xenopus l |
| 13 | 1078.2 | 25.8 | 2161 | 5 | AF055923 | AF055923 Xenopus l |
| 14 | 845.8 | 20.2 | 2011 | 5 | AF161270 | AF161270 Danio rer |
| c | 15 842.8 | 20.2 | 158105 | 2 | AC023375 | AC023375 Homo sapi |
| 16 | 820.4 | 19.6 | 2130 | 9 | AF117106 | AF117106 Homo sapi |
| 17 | 817.2 | 19.5 | 2381 | 9 | AF198254 | AF198254 Homo sapi |
| 18 | 814 | 19.5 | 2021 | 5 | AF026527 | AF026527 Gallus ga |
| 19 | 802.8 | 19.2 | 2223 | 10 | AF061569 | AF061569 Mus muscu |
| 20 | 710.4 | 17.0 | 2010 | 9 | AF117107 | AF117107 Homo sapi |
| 21 | 695.6 | 16.6 | 3667 | 9 | AF057352 | AF057352 Homo sapi |
| 22 | 658 | 15.7 | 2437 | 10 | M0SYBIC | L35549 Mus musculu |
| c | 23 627.4 | 15.0 | 59747 | 2 | AC069053 | AC069053 Homo sapi |
| 24 | 598.4 | 14.3 | 142971 | 9 | AC020629 | AC020629 Homo sapi |
| 25 | 590.2 | 14.1 | 147556 | 2 | AC011007 | AC011007 Homo sapi |
| 26 | 572.4 | 13.7 | 182695 | 2 | AC015706 | AC015706 Homo sapi |
| c | 27 572.4 | 13.7 | 184653 | 2 | AL596177 | AL596177 Homo sapi |
| 28 | 570.8 | 13.7 | 93682 | 2 | AL139152 | AL139152 Homo sapi |
| c | 29 554 | 13.3 | 57261 | 2 | AC090245 | AC090245 Homo sapi |
| 30 | 473.4 | 11.3 | 57261 | 2 | AC090245 | AC090245 Homo sapi |
| 31 | 427 | 10.2 | 165600 | 2 | AC079780 | AC079780 Homo sapi |
| c | 32 303.6 | 7.3 | 179501 | 2 | AC090107 | AC090107 Homo sapi |
| 33 | 286.2 | 6.8 | 62635 | 2 | AC087709 | AC087709 Homo sapi |
| c | 34 282 | 6.7 | 104668 | 9 | AC021876 | AC021876 Homo sapi |
| 35 | 226 | 5.4 | 51119 | 2 | AC090865 | AC090865 Homo sapi |
| c | 36 225.6 | 5.4 | 67823 | 2 | AC026579 | AC026579 Homo sapi |
| 37 | 213.8 | 5.1 | 67823 | 2 | AC026579 | AC026579 Homo sapi |
| 38 | 172.4 | 4.1 | 3188 | 3 | AF241237 | AF241237 Drosophill |
| 39 | 136.8 | 3.3 | 158105 | 2 | AC023375 | AC023375 Homo sapi |
| 40 | 122.4 | 2.9 | 258088 | 2 | AC073766 | AC073766 Mus muscu |
| 41 | 100.2 | 2.4 | 64442 | 2 | AC091133 | AC091133 Homo sapi |
| c | 42 100 | 2.4 | 159122 | 2 | AC025556 | AC025556 Homo sapi |
| c | 43 95.4 | 2.3 | 201739 | 2 | AC090288 | AC090288 Mus muscu |
| c | 44 94.8 | 2.3 | 239837 | 10 | AC084407 | AC084407 Mus Muscu |
| 45 | 93 | 2.2 | 46580 | 2 | AC012929 | AC012929 Drosophill |

ALIGNMENTS

| | | | | | | |
|------------|---|---------|------|-----|-------------|----------|
| RESULT | 1 | | | | | |
| LOCUS | HSU971188 | 4181 bp | mRNA | PRI | 11-SEP-1998 | |
| DEFINITION | Homo sapiens putative RNA binding protein KOC (koc) | | | | | complete |
| ACCESSION | U971188 | | | | | |
| VERSION | U971188.1 | | | | | |
| KEYWORDS | | | | | | |
| SOURCE | human. | | | | | |
| ORGANISM | Homo sapiens | | | | | |
| REFERENCE | | | | | | |
| AUTHORS | 1 (bases 1 to 4181) | | | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| | Zimmerhackl,F., Hameister,H., Varga,G., Friess,H., Buchler,M., | | | | | |
| | Beger,H.G., Villa,M.R., Adler,G. and Gress,T.M. | | | | | |
| TITLE | Cloning of a gene highly overexpressed in cancer coding for a novel | | | | | |
| JOURNAL | KH-domain containing protein | | | | | |
| MEDLINE | Oncogene 14 (22), 2729-2733 (1997) | | | | | |
| | 97322117 | | | | | |

Db 1501 |||||CGGTGCCATCATCGCAAGCGGCAGCACATCAAGCAGCTTCTCGCTTTCGTGGAGC 1560
Qy 1561 |||||ttcaattaaagtctccagcggaagcaccagatctaaagtgaagatggtgaattatcac 1620
Db 1561 |||||TTCAATTAAGATTGCTCCAGCGGAAGCACCAGATGCTAAAGTGAGATGGGTGATTATCAC 1620
Qy 1621 |||||tggaccaccagaggctcagttcaaggctcagggagaattttatggaaaaattaaagaaga 1680
Db 1621 |||||TGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAGAAAGA 1680
Qy 1681 |||||aaactttgttagtctaagaagaaggtgaaacttgaagctcata toagagtccatccett 1740
Db 1681 |||||AAACTTTGTTAGTCTCTAAAGAAAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCATCCCTT 1740
Qy 1741 |||||tgctctgagcagagttattggaaaaaggagcgaacaggtgaatgaacttcagaatttgc 1800
Db 1741 |||||TGCTCTGGCAGAGTTATTGGAAAAGGAGGCAAAACGGTGAATGAACTTTCAGAAATTTGTC 1800
Qy 1801 |||||aagtgcaagaattgttgcctctgtgaccagacacactgatgagaatgaccaagtgtgtg 1860
Db 1801 |||||AAGTGCAGAAAGTTGTTGCTCCCTCGTGACCAGACACCTGATGAGATGACCAAGTGGTTGT 1860
Qy 1861 |||||caaaataactgctcactctatgcttgcaggttgcaggttgcaggaataattcaggaattct 1920
Db 1861 |||||CAAAAATAACTGCTCACTTCTATGCTTGCCAGGTTGCCAGAGAAAATTCAGGAAAATTCCT 1920
Qy 1921 |||||gactcaggttaagcagcacaacaagaaggtcctgcaaaagtgaacacctcagtcacag 1980
Db 1921 |||||GACTCAGGTAAGAGCAGCAACACAGAGAGGCTCTGCAAAAGTGACCACCTCAGTCAAG 1980
Qy 1981 |||||acggaagtaaggtcaggaacagccaccacagaggcagatgccaacccaagaacaga 2040
Db 1981 |||||ACGGAAGTAAAGGCTCAGGAACAGCCACCACACAGAGGCGAGTGCCAACCAAGACAGA 2040
Qy 2041 |||||tbgtctaaacacagatgggctgaccccttatccagaaatcacatgcaagaattttac 2100
Db 2041 |||||TTGCTTAACCAACAGATGGGCGCTGACCCCTTATCCAGAAATCACATGCACAAGTTTATAC 2100
Qy 2101 |||||ctagccagttgttctgagaccaggaacttttgaacctcgtctctgtgagaatgtat 2160
Db 2101 |||||CTAGCCAGTTGTTCTTGAGGACCAGGCAACTTTTGAACCTCCTGCTGTGAGAAATGTAT 2160
Qy 2161 |||||actttatgctctctgaaatgtatgacaccagcttttaaacaacaaacaaacaaacaa 2220
Db 2161 |||||ACTTTATGCTCTCTGAAATGTATGACACCCAGCTTTAAACAAACAAACAAACAA 2220
Qy 2221 |||||aaaaggttgggagggaggaagaagagctctgcacttcccttctgttagtctca 2280
Db 2221 |||||AAAAGGTTGGGGAGGGGGAAGAGAGAGCTCTGCACCTTCCCTTTCTGTTAGTCTCA 2280
Qy 2281 |||||cagataacagatatcttaattctcttcttaattatcccccataatgccagaaattggctta 2340
Db 2281 |||||CAGTATAACAGATATTCTAAATTTCTTTTAATATCCCCCATATGCCAGAAAATGGCTTA 2340
Qy 2341 |||||atgatgctctcaataatcatcaatagattgctctcaatccaatgtttaaattgga 2400
Db 2341 |||||ATGATGCTTTTCACTAAATTCATAATAGATTGCTCTTAATCCCAATTTGTTAAATGGA 2400
Qy 2401 |||||tcagaataattatcacaggaacttaaatgttgaagccattagcatagaaaaactgttctca 2460
Db 2401 |||||TCAGAATAATTTATCAGGAACCTTAAATGTTAAGCCATTAGCATAGAAAACTGTTCTCA 2460
Qy 2461 |||||gtttatttttaccataacactaacatgagtaacctaaggaagtgctgaattggttggc 2520
Db 2461 |||||GTTTATTTTTTACCCTAACACTAACATGAGTAACCTTAAGGGAAGTGTGAATGGTTGGC 2520
Qy 2521 |||||aggggtattaaacgctattttactcaactaccctcaggtattcagtaacataagaaga 2580
Db 2521 |||||AGGGGTATTAAACGGTGCAATTTTACTCAACTACCTCAGGTATTCAGTAATAACAATGAAAA 2580
Qy 2581 |||||gcaaaaattgttccttttttggaaaattttatatactttataatgataagtcacaacq 2640
|||||

Db 2581 |||||GCAAAATTGTTCTCTTTTGTGAAAAATTTATATATACTTTATATGATAGAAGTCCAACCG 2640
Qy 2641 |||||ttttttaaaaaataaattttaaaatttaacagcaacttaacagcagcaataaataagatttt 2700
Db 2641 |||||TTTTTTAAAAATAAATTTAAAAATTTAACAGCANTCAGCTAACAGGCAAAATTAAGATT 2700
Qy 2701 |||||tacttctggctgggtgacagtaaaagctggaaaaataatttcagggtttttttgagggctttg 2760
Db 2701 |||||TACTTCTGGCTGGTGACAGTAAAGCTGGAAAAATTAATTTCAAGGCTTTTGTGAGGCTTTG 2760
Qy 2761 |||||acacagttattagtttaataacaaatgttcaaaaaatacggagcagtgctagtatctaggaga 2820
Db 2761 |||||ACACAGTTATTATTAGTTAAATCAAAATGTTCAAAAATACGGAGCAGTGCCTAGTATCTGGAGA 2820
Qy 2821 |||||gcagcaactaccattattcttccattatagttgggaagttttgacggtactacaaca 2880
Db 2821 |||||GCAGCACTACCATTTAATCTTCTTCAATTTAGTTGGGAAGTTTTCACGGTACTAACAAA 2880
Qy 2881 |||||gtggctgcaggagattttggaaaggctgggtttaaattgggttcaggagacttcagttttt 2940
Db 2881 |||||GTGGTGCAGGAGATTTTGGAACGGCTGGTTTAAATGGCTTCAGGAGACTTCAGTTT 2940
Qy 2941 |||||gtttagctacatgattgaatgataataaagctttgtgcttctgactatcaataacctaa 3000
Db 2941 |||||GTTTAGCTACATGATTGAATGCATAATAATGCTTTGTGCTCTGACTATCAATACCTAA 3000
Qy 3001 |||||agaagtgcatcagtgaaagatgcaagactttcaactgactgagcaaaagcaagcttta 3060
Db 3001 |||||AGAAAGTGCACTCAGTGAAGAGATGCAAGACTTTTCACTGACTGGCAAAAAGCAAGCTTTA 3060
Qy 3061 |||||gcttgccttataggatgcttagttggccactacacttcagaccaatgggagcagctataga 3120
Db 3061 |||||GCTTGTCTTATAGGATGCTTAGTTTGCCACTACACTTCAGACCAATGGGACAGTCAATAG 3120
Qy 3121 |||||tggttgacagtggtttaaagcgaacaaaggctacatttccatggggcagcagctgtcat 3180
Db 3121 |||||TGTGTGACAGTGTTTAAACGCAACAAAGGCTACATTTTCCATGGGGCCAGCACTGTCTAT 3180
Qy 3181 |||||gagcctcactaagctattttgaagatttttaagcactgataaattaaaaaataaaaaa 3240
Db 3181 |||||GAGCCTCCTAAGCTAATTTGAAGATTTTAAAGCATTTTAAAGCATGATAAATTTAAAAA 3240
Qy 3241 |||||aaattagactccacotctaagtagtaaaagtataacagagatttctgtactctgtgcaatcag 3300
Db 3241 |||||AAATTAGACTCCACCTTAAGTAGTAAAGTATAACAGAGTTTCTGTATCTGTGCAATCAG 3300
Qy 3301 |||||ttctttgaaaaaaaagtcacaaagatagagaatacacagaaaaagttttinggagataattg 3360
Db 3301 |||||TTCTTTGAAAAAAAAGTCAAAAGATAGAGAAATACAGAAAAAGTTTNGGGATATAATTTG 3360
Qy 3361 |||||aatgactgtaaaaacatatagaccttligataacgaactcatttgcactccttgacagca 3420
Db 3361 |||||AATGACTGTGAAAAACATATGACCTTTGATAACGAACCTCATTTGCTCACTCCTTGACAGCA 3420
Qy 3421 |||||aagccagtcagtcacaaattgtgtgggtggtgggttctccaagggccagcgtgctctctg 3480
Db 3421 |||||AAGCCAGTACGTACAAATTTGTTGGTGGTGGTGGTCTCCAAGGCCACGCTGCTCTCTG 3480
Qy 3481 |||||aatgtatttttgagtttggnttgnaaagatgatacagncatggttacactgatccttnaa 3540
Db 3481 |||||AATTGATTTTTTGGTTTTGGNTTGNAAAGATGATCACAGNCAATGTTACACTGATCTTNA 3540
Qy 3541 |||||ggaacatatntataacccctttaaaaaataatccccctgctcactcttatttttgagatga 3600
Db 3541 |||||GGACATATTNTTAAACCCCTTTAAAAAAAATCCCTGCTCAATCTTATTTATTCGAGATGA 3600
Qy 3601 |||||atttcagacagactagatgcttctgaagatcaattagacattntgaaaaatgatttaa 3660
Db 3601 |||||ATTTCCATACAGACTAGATGCTTCTTGAAGATCAATTAGACATTTGAAAAATGATTTAA 3660
Qy 3661 |||||agtgtttctcctaagtctctgaaaaacaagtttctttttagtttttaacccaacaaagt 3720
Db 3661 |||||AGTGTGTTTCCTTAATGTTCTCTGAAAAACAAGTTTCTTTTGTAGTTTAAACCAAAAAGTG 3720

Db 901 |||||CATTCGGAAACATCACCAACAGACCAGCTCAAAATCGATGTCACCGTAAGAAAAATGC 960
 Qy 961 999gctgtcagaaagtactactctctactctctactctctgaagcaacctctcgcgcttg 1020
 Db 961 GGGGCTGCTGAGAAGTCGATTACTATCTCTCTACTCTCTGAAGGCACCTCTCGGCGCTG 1020
 Qy 1021 taagtctattctgagagattatgcataaagaagctcaagatatataaataatccagaagagat 1080
 Db 1021 TAAGTCTATTCTGGAGATTATGCATAAAGAAAGCTCAAGATATAAATTCACAGAAAGAT 1080
 Qy 1081 cccctgaagatttagctcataataactttagtggacgtcttattgtaagaaggag 1140
 Db 1081 |||||CCCCCTGAAGATTTAGTCTATAAATCTTGTGGAGCTCTTATGTAAGAAGGAAG 1140
 Qy 1141 aaatctaaaaaattgagcaagacacagacactaaaatcacgatatctccattgcagga 1200
 Db 1141 AAATCTTAAAAAATTGAGCAAGACACAGACATAAATACAGTATCTCCATTCACAGA 1200
 Qy 1201 attgagctgtataatccagaaacgcactattacaggttaaaggcaattgttgacatgtgc 1260
 Db 1201 ATTGACGCTGTATAATCCAGAAGCACATATACAGTTAAAGCAATGTTGAGACATGTGC 1260
 Qy 1261 caaagctgagagagatcatgaagaaatcaggagctcttatgaaaaatgatattgcttc 1320
 Db 1261 CAAAGCTGAGGAGGAGATCATCAAGAAATCAGGAGTCTTATGAAAAATGATATTGCTTC 1320
 Qy 1321 tatgaatctcaagcacatttaattccctggattaaatctgaacgcttggctgtctccc 1380
 Db 1321 TATGAACTCTCAAGCAATTTAAATCTCTGGATTAATCTGAACGCCCTGGGCTCTGTTCCC 1380
 Qy 1381 acccaactcaggagatgccactccacactcaggccagggcccttcagcatgaactccctccta 1440
 Db 1381 ACCCACTTCAGGATGCCACTCCCACTCAGGGCCCCCTTCAGCCATGACTCCTCCCTA 1440
 Qy 1441 cccgcagttttagcaatcagaacagagactgttcatcagtttatccacagctctatcagtt 1500
 Db 1441 CCCGCAGTTTGAGCAATCAGAACCGAGACTGTTCAATCAGTTTATCCAGCTCTATCAGT 1500
 Qy 1501 cgggtccatcatcggcagcaggccagcacatcaagcagcttctcgtcttgctggagc 1560
 Db 1501 CGGTGCCATCATCGCAAGAGGGCCAGCACATCAAGCAGCTTCTCGCTTTGCTGGAGC 1560
 Qy 1561 ttcaattaagattgtccagcggagcagccagatgtctaagtggagtgagttatcac 1620
 Db 1561 TTCAATTAAATGCTCCAGCGGAAGCACAGATGCTTAAAGTGAGATGGTGAATTATCAC 1620
 Qy 1621 tggaccaccagagctcagttcagggtcagggaagaatttatggaaaaattaaagaaga 1680
 Db 1621 TGGACCAACAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAGAAGA 1680
 Qy 1681 aaactttgttagtctaagaagagtgtaaaacttgaagctcatatcagagtgccatcctt 1740
 Db 1681 AAACTTTGTTAGTCTCTAAAGAAGAGGTGAAGCTTTGAAGCTCATATFCAGAGTGCCATCCTT 1740
 Qy 1741 tgcgtcggcagagttattggaaaaggaggcaaaacggtgaatgaacttcagaaattgtc 1800
 Db 1741 TGCCTGCGCAGAGTTATTGGAAAAGGAGGCAAAACGGTGAATGCAGAAATTTGTC 1800
 Qy 1801 aagtcagaagtgtgttcctcgtgaccagacacacctgatgagaatgaccgaagtgtgtg 1860
 Db 1801 AAGTGCAGAAGTTGTTGTCCTCGTGACCAGACACCTGATGAGATGACCAAGTGGTGT 1860
 Qy 1861 caaaaactagtcacttctatgttgccaggttgcccagagaaaaattcaggaaattct 1920
 Db 1861 CAAAATAACTGGTCACTTCTATGCTTGCCAGGTGCCCCAGAGAAAAATTCAGGAAATTC 1920
 Qy 1921 gactcaggttaagcagcaccacaacagaggtctctgcaaaagtgaccacacctcagtcag 1980
 Db 1921 GACTCAGGTAAAGCAGCACCAACACAGAGGCTCTGCAAGTGGACCACTCAGTCAAG 1980
 Qy 1981 acggaagtaaaggctcaggaacagcccccacacagagagcagatgccaacccaaagacaga 2040
 Db |||||

Db 1981 ACGGAAGTAAAGGCTCAGGAACAGCCCCACCACAGAGGCGAGATGCCAAACCAAGACAGA 2040
 Qy 2041 ttgcttaacaaacagatggcgctgaccccttatccagaaatcacatgcaagaagttttac 2100
 Db 2041 TTGCTTAACCAACAGATGGCGCTGACCCCTTATCCAGAATCACATGACAAAGTTTTTAC 2100
 Qy 2101 ctagccagttgtttctgaggaaccaggcaactttgaactcctctctctctgagaaatgat 2160
 Db 2101 CTAGCCAGTTGTTTCTGAGGACAGGCAACTTTTGAACCTCTCTCTCTGAGAAATGTAT 2160
 Qy 2161 acttattctctctgaatgtatgacaccagcttttaaacaacaaacaaacaaacaaacaa 2220
 Db 2161 ACTTTATGCTCTCTGAATGTATGACACCCAGCTTTAAACAAACAAACAAACAAACAA 2220
 Qy 2221 aaaaaggtggggagggaggaagagagctctgcacttccctttgttagtctca 2280
 Db 2221 AAAAGGTTGGGAGGAGGAGAAAGAGAGCTCTGCACCTTCCCTTTGTTGTAGTCTCA 2280
 Qy 2281 cagtaaacagatatcttaattcttcttaatttcccccaataatgccagaaaaattggctta 2340
 Db 2281 CAGTATAACAGATATTCTAAATCTTCTTAAATTCCTCCCATATATGCCAGAAATGGCTTA 2340
 Qy 2341 atgatgcttcaactaaattcatcaaaatagattgctccttaaatcccaattgttaaaattgga 2400
 Db 2341 ATGATGCTTTCACTAAATTCATCAAAATAGATTGCTCTAAATCCAATTTGTTAAATTTGA 2400
 Qy 2401 tcagaaataattacaggaacttaattgttaagcatttagcatagaaaaactgttctca 2460
 Db 2401 TCAGAAATAATTATCAGAGAACTTAAATGTTAAGCCATTTAGCATAGAAAACTGTTCTCA 2460
 Qy 2461 gttttatttttaacctaacactaacatgagtaacctaaaggaagtgcgtgaattgggttgc 2520
 Db 2461 GTTTTATTATTTACCTAAACACTAACATGAGTAACCTAAGGGAAGTGTGAATGGTGTGC 2520
 Qy 2521 aggggtattaaacgctgtttttactcaactcaactcagggtattcagtaatacaatgaaaa 2580
 Db 2521 AGGGGTATTAAACGTCATTTTTTACTCAACTACCTCAGGTATTCAGTAATACAAATGAAA 2580
 Qy 2581 gcaaaatgtctcttttttgaaaatttatatactttatataatgatagaagcccaacg 2640
 Db 2581 GCAAAATTTGTTCTTTTTTTCAAAAATTTATATACTTTATATGATAGAAAGTCCAAACG 2640
 Qy 2641 ttttttaaaaaataaatttaaatttaacagcaatcagctaacacagcagcaaatlaagatttt 2700
 Db 2641 TTTTTTAAAAATAAATTTAAATTTAACAGCAATCAGCTAACAGGCAAAATTAAGATTTT 2700
 Qy 2701 tactctcgtcgtgagcaagaagctggaataatlaatttcagggtttttttgagggctttg 2760
 Db 2701 TACTTCTGGCTGGTGACAGTAAAGCTGGAAAAATTAATTTTCAGGGTTTTTTTGAGGCTTTTG 2760
 Qy 2761 acacagttattagttaaatcaaatgttcaaaaatcaggagcagtgctagtatctgagga 2820
 Db 2761 ACACAGTTATTATAGTTAAATCAAAATGTTCAAAAATACGGAGCAGTGCCTAGTACTCGAGA 2820
 Qy 2821 ggcagcactaccatttattcttttcattttagttggaaaagtttttgcggtgactaaacaa 2880
 Db 2821 GCAGCAGTACCATTTATTTCTTTCAITTTATAGTTGGGAAGTTTTCACGGTACTACAAA 2880
 Qy 2881 gtggtcgcagagagatttggaaacgctggtttaaatggttcaggagagactcagtttttt 2940
 Db 2881 GTGTCGCGAGGAGATTTTGGAAACGGCTGTTTTAAATGGCTTCAGGAGACTTCAGTTTTTT 2940
 Qy 2941 gtttagctacatgattgaatgataataatgtttgtcttctgaactatcaataactaa 3000
 Db 2941 GTTTAGCTACATGATTGAATGATCAATAAATGCTTTGTGCTCTTGACTACTCAATACCTAA 3000
 Qy 3001 agaaagtgcacagtgaaagagatgcaagactttcaactgactggcaaaaagcaagcttta 3060
 Db 3001 AGAAAGTCATCAGTCAAGAGATGCAAGACTTTCACCTGACTGCGCAAAAAGCAAGCTTTA 3060
 Qy 3061 gcttcttattaggatgcttagtttggccactacacttcagaccaaattgggacagtcataga 3120
 Db 3061 GCTTGTCTTATAGGATGCTTAGTTTGCCACTACACTTCAGACCAATTCAGGACAGTCATAGA 3120

| | | | |
|----|------|---|------|
| Db | 197 | TCATACCGTGGGTGTTTTAATTTTCGTTTTTCGTTATTTTTTTTAAACAACC | 256 |
| Qy | 241 | actcttcaaatgaacaaactgtatatcgaaacctcagcgagaaacgcgcgccttcgga | 300 |
| Db | 257 | ACTCTTCACAATCAACAACTGTATATCGGAAACCTCAGCGAGAACGCGGCCCTCGGA | 316 |
| Qy | 301 | cctagaagaatctctcaagagcgaagatcccgtgtctgggaaccttccttgtagaac | 360 |
| Db | 317 | CCTAGAAAGTATCTTCAAGGAGCCGAAGATCCGGTGTGCGGACCTTCTGTGTGAAGAC | 376 |
| Qy | 361 | tggctaagcgtctgtgaactgcccgaagagcttgggccctcaaggccatcgagcgct | 420 |
| Db | 377 | TGGCTACGCTCTCGTGGACTGCCCCGAGAGAGCTTGGCCCTCAAGGCCATCGAGGCGCT | 436 |
| Qy | 421 | ttcaggtcaaaatagaactcacggaagaccccatagaagtgtgacactcgtgcccaaaaag | 480 |
| Db | 437 | TTCAGGTAAATAGAACTGCACGGGAACCCATAGAAATTGAGGACACTCGGTCCCAAAAG | 496 |
| Qy | 481 | gcaaggattcggaaacttcagatacagaataatcccgcctcatctacagtggaggtgct | 540 |
| Db | 497 | GCAAGGATTTCGAAACTTCAGATACGAAATATCCGCCCTCATTTACAGTGGGAGGTGCT | 556 |
| Qy | 541 | ggatagtttactagttccagataggagtggagtgtagcaagtgaacctgactc | 600 |
| Db | 557 | GGATAGTPTTACTASTCCAGTATGGAGTGGTGAGAGCTGTGAGCAAGTGAACTGACTC | 616 |
| Qy | 601 | ggaaactcgagttgaaatgaacctattccagtaagacacaaactagacaacactaga | 660 |
| Db | 617 | GGAAACTGCAGTTGTAAATGTAACTATTCCAGTAAGACCACACTAGCAAGCACATAGA | 676 |
| Qy | 661 | caactgaatggatttcagttagagaatttcaccttgaagtgaagctcatatccctgatga | 720 |
| Db | 677 | CAAACTGAATGGATTTTCAGTTAGAGAAATTACACTTGAAGTAGCCCTATATCCTGTATGA | 736 |
| Qy | 721 | aatggcccagcaaaaaaccccttgacgaagccccgaggtgcgcggggggtcttgggcaag | 780 |
| Db | 737 | AATGGCCGCCACGACAAACCCCTTGACGAGCCCCGAGGTCGCCGGGGGCTTGGGCAGAG | 796 |
| Qy | 781 | ggctcctcaagcaggggtctccaggatccgataccaagcagaaaaacatgtgattgcc | 840 |
| Db | 797 | GGGCTCTCAAGCAGGGGTCTCCAGATPCGATATCCAAGCAGAAACCATGTGATTTGCC | 856 |
| Qy | 841 | ctgcgcctgctgtttcccacccaaattgttgagcccatcatgaaaagaagtgccac | 900 |
| Db | 857 | CTGCGCCTCTGTGTTCCACCCTTGTGTGGAGCCATCATAGGAAAGAAGGTGCCAC | 916 |
| Qy | 901 | cattcgaaacatcaccaaacagaccagtcctaaaaatcgatgtccaccgtaaagaaatgc | 960 |
| Db | 917 | CATTTCGGAACATCACCAAAACAGACCCAGTCTAAAAATCGATGTCCACCGCTAAAGAAAAATGC | 976 |
| Qy | 961 | ggggctctcgaagatcgattactatcctctactcctgaagccacctctcgcgcttg | 1020 |
| Db | 977 | GGGGCTCTCGAAGTCGATTACTATCTCTCTACTCCTGAAGGCACCTCTCGCGGCTTG | 1036 |
| Qy | 1021 | taagtctattctggagattatgcaagaagactcaagataataaaattcacagaagaagat | 1080 |
| Db | 1037 | TAAGTCTATTTCGAGATTATGCATAAGGAAGCTCAAGATATAAATTCACAGAAAGAT | 1096 |
| Qy | 1081 | cccttgaagattttagtcataataactttgttgagcgtcttattggtaaagaaggaag | 1140 |
| Db | 1097 | CCCTTTGAAGATTTTAGTCTAATAAATCTTTGTTGGAGCTCTTATTGTTAAAGAAGGAAG | 1156 |
| Qy | 1141 | aaatcttaaaaaaattgacaagacacagacactaaaaatcacgatatctccattcgagg | 1200 |
| Db | 1157 | AAATCTTAAAAAAATTTGACCAAGACACAGACACTAAAAATACGATATCTCCATTTGCAGGA | 1216 |
| Qy | 1201 | attgaactgtataatccagaacgcaactattacagttaaaggcaatgttgagacatgtc | 1260 |
| Db | 1217 | ATTGACGCTGTATTAATCCAGACGCNACTATTACAGTTTAAAGGCNAATGTTGAGACATGTC | 1276 |
| Qy | 1261 | caagctgaggaggagatcatgaagaaatcaggaggtctcttatgaaatgatattgcttc | 1320 |

| | | | |
|----|------|--|------|
| Db | 1277 | CAAACTGAGGAGGATCATGAAGAAAATCAGGGAGTCTTATGAANAATGATATTCCTTC | 1330 |
| Qy | 1321 | tatgaatctcaagcacatttaattcttgattaaacttgaaagccttggtctgttccc | 1380 |
| Db | 1337 | TATGAATCTTCAAGCACATTTAATTCCTGGAATTAATCTGAAGCCTTGGTCTGTCTCC | 1396 |
| Qy | 1381 | accacttcagggaatgcacacctcccaacctcaggggcccccttcagccaatgactcctcccta | 1440 |
| Db | 1397 | ACCACACTTCAGGGATGCCACCTCCCACTTCAGGGCCCCCTTCAGCCATGACTCCTCCCTA | 1456 |
| Qy | 1441 | ccgcagatttgcaatcagaacaacggagactgttcaactgattatcccagctctatcagt | 1500 |
| Db | 1457 | CCGCAGATTTGAGCAATCAGAAACGGAGACTGTTCATCTGTTTATCCCACTCTATCAGT | 1516 |
| Qy | 1501 | cqgtgccatcatcgccaagcaggccagcacatcaagcagctttctcgttttcttgagac | 1560 |
| Db | 1517 | CGGTGCCATCATCGCAACAGGGCCAGCACATCAAGCAGCTTCTCGCTTTGCTGGAGC | 1576 |
| Qy | 1561 | ttaattaagaatgtctccagcgggaagcaccagatgctaagtgaggatggtgattatcac | 1620 |
| Db | 1577 | TTCAATTAAATTTGCTCCAGCAGAAAGCACCATGCTAAAGTGAGGATGGTGATTATCAC | 1636 |
| Qy | 1621 | tgaccaccagaggctcagttcaaggctcagggaagaatttatggaaaaattaaagaaga | 1680 |
| Db | 1637 | TGGACCAACAGAGGCTCAGTTCAAGGCTCAGGGAAGAAATTTATGSAANAATTTAAAGAAGA | 1696 |
| Qy | 1681 | aaactttgtagtcctaaagagaggtgaacttgaagctcatatcagagtgccatcctt | 1740 |
| Db | 1697 | AAACTTTGTAGTCTTAAGAGAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTT | 1756 |
| Qy | 1741 | tgctctggcagagttattggaaaaaggaggaacacggtgaatgaacttcagaatttgc | 1800 |
| Db | 1757 | TGCTGCTGSCACAGTTATTGGAAAAGGAGGCAAAACGGTGAATGAACCTCAGAAATTTGTC | 1816 |
| Qy | 1801 | aagtgcagaagtgttgtccctgtgaccagacacctgatgagaatgcacaaagtgttgt | 1860 |
| Db | 1817 | AAGTCGAGAAGTTGTTGTCCCTCGTGACCAGACACCTGATGAGAATGACCAAGTGTGTG | 1876 |
| Qy | 1861 | caaaataactgctcacttctatgcttgccaggttgccagagaagaaatcagaaattct | 1920 |
| Db | 1877 | CAAAATAACTGGTCACTTCTATGCTTGCCAGGTTGCCAGAGAAAATTCAGGAATTTCT | 1936 |
| Qy | 1921 | gactcaggttaagcagcagcccaacaacagaaggtctctgcaaaagtggaccacctcagtcag | 1980 |
| Db | 1937 | GACTCAGGTAAAGCAGCACCAACACAGAAGGCTCTGCAAGGTGGACCACTTCAGTCAAG | 1996 |
| Qy | 1981 | acggaagtaaaaggttcaggaacaacagcccaccacagaggcagatgccaaaccaagaacaga | 2040 |
| Db | 1997 | ACGGAAGTAAAGGCTCAGAAAACAGCCACACAGAGGCAGATGCCAAACCAAGACAGA | 2056 |
| Qy | 2041 | tgtcttaaccaacagatggcgtgaacccctatccagaatcacatgcacaagtgtttac | 2100 |
| Db | 2057 | TTGCTTAACCAACAGATGGCGCTGACCCCTTATCCAGAAATCATATGCACAGTTTTTTAC | 2116 |
| Qy | 2101 | ctagccagttgtttctgagcacaggcaactttgaaactcctgtctctgtgagaatgtat | 2160 |
| Db | 2117 | CTAGCCAGTTGTTCTTGAGGACCAGGCAACTTTTGAATCTCTGTCTCTGTGGAATGTAT | 2176 |
| Qy | 2161 | actttatgctctgaaatgtatgacaccacagcttttaaacaacaacaacaacaacaaa | 2220 |
| Db | 2177 | ACTTTATGCTCTGAAAATGTATGACACCCAGCTTTAAACAAACAAACAAACAAA | 2236 |
| Qy | 2221 | aaaagggtggggaggagggaagagaagactctgcacttcccttgttgtagttctca | 2280 |
| Db | 2237 | AAAAGGTTGGGGAGGGGAAAAGAGAGGCTCTGCACTTCCCTTTGTTGTAGTCTCA | 2296 |
| Qy | 2281 | cagttataacagatattctaatcttcttaatttcccccatatgcagaataatggcttta | 2340 |
| Db | 2297 | CAGTATAACAGATATCTTAATCTCTTAAATATCCCCCAATAATGCCAGAAATTGGCTTA | 2356 |
| Qy | 2341 | atgagtctttcaactaaattcataatagattgctcctcaatccaattgttaaaattgga | 2400 |
| Db | 2357 | ATGAGTCTTTCACTAAATTCATCAATAGATGCTCCTAAATCCAATGTTTAAATTTGGA | 2416 |

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 497J21. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 427A4 (298049) is at 129333 in this sequence.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

Assembly was confirmed by BamHI and HindIII digests. Discrepancy noted in EcoRI digest.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

497J21 is from the library RBC13 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see [http://bacpac.med.buffalo.edu/ VECTOR: pcypac2](http://bacpac.med.buffalo.edu/VECTOR: pcypac2).

FEATURES

| | source |
|---------------|--|
| | 1. 129432 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /chromosome="6" |
| | /map="q26-27" |
| | /clone="RF3-497J21" |
| | /clone_lib="RPCI-3" |
| | complement(1..62) |
| repeat_region | /note="AluX repeat: matches 62..1 of consensus" |
| misc_feature | 27..424 |
| repeat_region | /note="match: STS AL022447" |
| repeat_region | 63..231 |
| repeat_region | /note="MER20 repeat: matches 53..218 of consensus" |
| repeat_region | 433..591 |
| repeat_region | /note="3 copies 53 mer 86% conserved" |
| repeat_region | complement(1513..1915) |
| repeat_region | /note="MLT1B repeat: matches 382..2 of consensus" |
| repeat_region | complement(2671..2713) |
| repeat_region | /note="MLT1B repeat: matches 47..3 of consensus" |
| repeat_region | complement(3291..3586) |
| repeat_region | /note="AluSq repeat: matches 298..1 of consensus" |
| repeat_region | 4989..5126 |
| repeat_region | /note="L1MB6 repeat: matches 1..141 of consensus" |
| repeat_region | 5137..5648 |
| repeat_region | /note="L1MB5 repeat: matches 382..923 of consensus" |
| repeat_region | complement(7362..7463) |
| repeat_region | /note="MIR repeat: matches 194..87 of consensus" |
| repeat_region | 8372..8668 |
| repeat_region | /note="AluX repeat: matches 2..300 of consensus" |
| repeat_region | 8670..8703 |
| repeat_region | /note="17 copies 2 mer aa 82% conserved" |
| repeat_region | 14252..14295 |
| repeat_region | /note="22 copies 2 mer gt 98% conserved" |
| repeat_region | 14254..14301 |
| repeat_region | /note="3 copies 16 mer 94% conserved" |
| repeat_region | 14254..14297 |
| repeat_region | /note="11 copies 4 mer gtgt 98% conserved" |
| repeat_region | 14441..14481 |
| repeat_region | /note="MIR2 repeat: matches 99..139 of consensus" |
| repeat_region | complement(16259..16324) |
| repeat_region | /note="MIR repeat: matches 157..92 of consensus" |
| repeat_region | complement(18532..18711) |
| repeat_region | /note="MIR repeat: matches 218..29 of consensus" |
| repeat_region | 20727..21002 |
| repeat_region | /note="AluX repeat: matches 2..296 of consensus" |
| repeat_region | 22965..23004 |
| repeat_region | /note="10 copies 4 mer caca 83% conserved" |
| repeat_region | complement(24989..25171) |
| | /note="L1MB3 repeat: matches 931..738 of consensus" |
| | complement(25204..25504) |
| | /note="AluSp repeat: matches 303..1 of consensus" |
| | complement(25761..26017) |
| | /note="AluY repeat: matches 301..45 of consensus" |
| | complement(27102..27194) |
| | /note="MIR repeat: matches 142..50 of consensus" |
| | 30090..30204 |
| | /note="5 copies 23 mer 94% conserved" |
| | complement(31383..31419) |
| | /note="MIR2 repeat: matches 113..76 of consensus" |
| | 31912..32368 |
| | /note="match: STS S53188; part 1 of genomic marker D6S193" |
| | 32276..33267 |
| | /note="31 copies 32 mer 91% conserved" |
| | 33141 |
| | /note="tandem repeat" |
| | 33204 |
| | /note="tandem repeat" |
| | 33222..33396 |
| | /note="tandem repeat and single clone region" |
| | 33268..33555 |
| | /note="9 copies 32 mer 88% conserved" |
| | 33288..33539 |
| | /note="17 copies 16 mer 89% conserved" |
| | 33473..33793 |
| | /note="match: STS S53190; part 2 of genomic marker D6S193" |
| | complement(34751..35259) |
| | /note="MER21B repeat: matches 754..240 of consensus" |
| | complement(35356..35625) |
| | /note="MER21B repeat: matches 282..5 of consensus" |
| | 36446..36618 |
| | /note="MLT1B repeat: matches 1..179 of consensus" |
| | 36619..36921 |
| | /note="AluSq repeat: matches 1..303 of consensus" |
| | 36927..37136 |
| | /note="MLT1B repeat: matches 175..390 of consensus" |
| | 38179..38310 |
| | /note="2 copies 66 mer 92% conserved" |
| | 38187..38318 |
| | /note="4 copies 33 mer 91% conserved" |
| | complement(39918..40225) |
| | /note="AluX repeat: matches 302..1 of consensus" |
| | complement(41241..41409) |
| | /note="L1 repeat: matches 5390..5208 of consensus" |
| | 42029..42091 |
| | /note="MIR2 repeat: matches 69..138 of consensus" |
| | complement(43307..43755) |
| | /note="match: GSS B76271" |
| | 46203..46713 |
| | /note="MER1A repeat: matches 1..527 of consensus" |
| | complement(50808..51108) |
| | /note="AluSp repeat: matches 303..2 of consensus" |
| | 52433..52723 |
| | /note="AluDb repeat: matches 5..301 of consensus" |
| | 53862..54166 |
| | /note="MER2 repeat: matches 1..324 of consensus" |
| | 54319..54430 |
| | /note="7 copies 16 mer 81% conserved" |
| | 54322..54417 |
| | /note="3 copies 32 mer 89% conserved" |
| | 54720..54767 |
| | /note="24 copies 2 mer tg 88% conserved" |
| | 58071..58370 |
| | /note="AluX repeat: matches 1..301 of consensus" |
| | 58587..58738 |
| | /note="4 copies 38 mer 86% conserved" |
| | complement(59963..60560) |
| | /note="match: GSS AQ001508" |
| | 60205..60273 |
| | /note="3 copies 23 mer 84% conserved" |
| | 61253..61299 |
| | /note="MIR repeat: matches 104..150 of consensus" |


```
Qy 731 cagcaaaaccccttgagcagcccgaggtgcgcggggcttgaggcagaggggctcctca 790
Db 481 CAGCAAAATCCCTCGCCACAGCTCCGGGGCGCGGGGCCAGCGGGGCTCATCC 540
Qy 791 aggcagggtctccagatccgtatccaagcagaaacacatgtgattgctctgcgcgtg 850
Db 541 AGCAGCGCTCTCAGGATCGGTCTCCAAGCAGAAAACCTGTGACCTGCCACTGCGCTG 600
Qy 851 ctggttcccccaaatattgttgagccatcataggaagaaagagtgccacattcggaa 910
Db 601 CTGGTCCCAACCCAGTTTGTGTGGACCATATAGAAAAAGAGTGCTACCATTCGCAAC 660
Qy 911 atcaaaaaacagaccagtcataaatcgatgtccacogtaaaagaaatgaggggctgct 970
Db 661 ATCAAAAGCAGACCCAGTCTAAATCGATGTCATCGTAAGGAGAAATACAGGGCCGCG 720
Qy 971 gagaagtcgattactatccctctactcctgaagccctctgcggtctgtaagctctatt 1030
Db 721 GAGAAGTCCATTACTATCTCTTACCCCTGAAGGCACCTCTCGCAGGTTGTAAGTCTATT 780
Qy 1031 ctggagattatgcataaggaagctcaagatatataaaattcacagaagagatccccctgaag 1090
Db 781 CTGGAGATTATGCATAAGGAAGCTCAAGATATAAAATTCACAGAGGAGATTCCCTTCAAG 840
Qy 1091 attttagctcataaactattgttgagcgtcttattgttaaaagaaagaaatctttaa 1150
Db 841 ATCTTAGCTCACATAACTTTGTAGGCGCTCTCATTTGTTAAAGAAAGAAACCTTTAA 900
Qy 1151 aaaattgagaagacagacactaaaatcacgatattctcattgcaggaattgacgtg 1210
Db 901 AAATCGAGCAAGACCGGACACTAAATCAATATCTTCNTTGCAGGAACGTGACCGTG 960
Qy 1211 tataatccagaagcactatatacagttaaaggaatgttgagacatgtgcaaaagctgag 1270
Db 961 TACAATCCGAACGCACCATTACAGTGAAGGCGAGTGTTCAGACGTGTGCGCAAGCGCGAG 1020
Qy 1271 gaggagatcatgaagaaatacaggagcttattgaataatgatattgtcttatgaattctt 1330
Db 1021 GAGGAAATATGAAGAGATGAGGAGCTCTATGAAATATATATTTGCTTCATGAATCTT 1080
Qy 1331 caagcacatttaattcctgattaaactgaagccttggtgtctgtccaccacactca 1390
Db 1081 CAAGCACATTTAATCCCTGATTAATCTGAATGCCCTTGGGTCTGTTCCACCACCGTCA 1140
Qy 1391 gggatgccactcccaacctcaggggcccttcagccaatgactctcctccctcccgagctt 1450
Db 1141 GGGATGCCACTCCCACTCAGGGCCCTCTCAACCCCTGACTCTCCTTACCACACAATTT 1200
Qy 1451 gagcaatcagaagcagactgttcatcagtttatccagctctatcagtcggtgcatc 1510
Db 1201 GAGCAATCAGAGACGGAGACTGTGCATCTGTTTATTTCCCGCCCTGTCCGTGGCGCCATC 1260
Qy 1511 atcggaagcagggccagcacatcaagcagcttctcgtcttctgagagctcaattaaag 1570
Db 1261 ATTGGCAAGCAGGGCCCAACATCAACACCTTCTCGCTTTGGGGAGCTTCGATTAAAG 1320
Qy 1571 attgctcagcgaagcaccagatgctaaagtggagtgatgattatcactggaccacca 1630
Db 1321 ATCGCTCCAGCGAAGCACCAGATGCTFAAAGTGGGATGCTGATTATCACTGGACCA 1380
Qy 1631 gaggctcaagtcaggctcagggaagaatttatggaataattaaagaagaaactttggtt 1690
Db 1381 GAGGCTCAGTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAGAAGAAAAAATTTGTT 1440
Qy 1691 agtctaaagaagagtgaaactgaagctcatatcagagtgccatcttctgctgctg 1750
Db 1441 AGTCTTAAAGAGAGGTGAACCTTGAAGCTCACATCAGAGTGGCGCTCTTGTCTGTGGC 1500
Qy 1751 agagttattggaagagggcaaaacggtgaatgaacttcagaaatttgcagaatgagaa 1810
Db 1501 AGAGTTATTGGGAAAGAGGCAAAACGGTGAATGAGCTCCAGAGTTTATCAAGTGTGAA 1560
Qy 1811 gttgtgtccctcgtgaccagacacactgatagagaatgaccaagtgggtgtgcaaaataact 1870
```

```
Db 1561 GTTGTCTCCCTGACACACACCTGATGAGATGATCAAGTAGTGTCAAAATAACT 1620
Qy 1871 ggtcaattctatgctgccaggttgccagagaaaaattcaggaaattcttgaactcagta 1930
Db 1621 GGCCACTTCTATGCTTGCCAGGTTGCCAGAGAAATTCAGAAATTTCTGACTCAGTA 1680
Qy 1931 aagcagcaccaacaacagaagggctctgaaagtggaccacctcagtcgaagcaggaagtaa 1990
Db 1681 AAGCAGCACCAGCAGCAAGAGCTCTGCAGAGTGGACCACCTCAGTCGAAGCGGAAGTAA 1740
```

RESULT 7

AC092447/c

LOCUS

DEFINITION

AC092447

AC092447

AC092447.1

HTG; HTGS_PHASE1

human

ORGANISM

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Waterston, R.H.

Unpublished

2 (bases 1 to 181369)

Waterston, R.H.

Direct Submission

Submitted (04-JUL-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

Project Information

Center project name: H.NH0760D02

NOTE: This is a 'working draft' sequence. It currently

consists of 30 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1170: contig of 1170 bp in length

1171: gap of unknown length

1271: contig of 1387 bp in length

2558: gap of unknown length

2757: contig of 1770 bp in length

4527: gap of unknown length

4628: contig of 1068 bp in length

5695: gap of unknown length

5795: contig of 1040 bp in length

6835: gap of unknown length

6935: contig of 2001 bp in length

8936: gap of unknown length

9037: contig of 1787 bp in length

10824: gap of unknown length

10924: contig of 3048 bp in length

13971: gap of unknown length

14071: contig of 3019 bp in length

17090: gap of unknown length

17190: contig of 2530 bp in length

17191: gap of unknown length

19720: contig of 3169 bp in length

19821: gap of unknown length

22989: contig of 5098 bp in length

23090: gap of unknown length

28187: contig of 5098 bp in length

| | | | |
|----|--------|--|---------|
| Db | 121631 | TAAAGAGAAATTTTGTAGTCTCTAAAGAGAGGTTAAACTTTGAAGCTTATATACAGT | 1215757 |
| Qy | 1732 | gccatccttctgctgagagatttggaaaagaggcacaacccggtgaatgaacttca | 1791 |
| Db | 121571 | GTATCTCTTCTGCTGGCGGAGTTATTGAAAAAGGAGGCAACCGTGAAATAACTTCA | 121512 |
| Qy | 1792 | gaattgtcgaagtgcagaagtgtgtccctcgtgaccagacacctgatgaaatgacca | 1851 |
| Db | 121511 | GAATTTGTCAAGTGCAGAAGTTGTGTCCCTCATGACCAGACACCTGATGAAATGACCA | 121452 |
| Qy | 1852 | agtgtgtgtaaaataactcgtgcaactctatgcttgcagagttgcccagagaaaaattca | 1911 |
| Db | 121451 | AGTGGTTGTCAAAATAACTGGTCACTTCTATGCTGTGCCAGTTGCTCAGAGAAAAATTCA | 121392 |
| Qy | 1912 | ggaattctgactcaggtcaagcagcaccaacaacagaaggctctcgaagtggaccacc | 1971 |
| Db | 121391 | GGCAATTTCTGACTCAGGTAAACACAGCA-CAACAGCAGAAGGCTCTGCAAGTGGACCACC | 121333 |
| Qy | 1972 | tcagtccaagcgggaagttaaaggctcaggaagaaacgcccacacagagcgcagatgcacaaacc | 2031 |
| Db | 121332 | TCAGTCAAGATGGAAGTAAAGACTCAGAGAAACAGCCACCACAGAGCGAGATGGCAAAACA | 121273 |
| Qy | 2032 | aaagacagattgcttaaccaaacagatgaggcgctgacccccctatccagaatcacatgcaca | 2091 |
| Db | 121272 | AAAGACAGATTGCTTTAACCAACAGACAGCGCGCTGACCCCTATCCAGAAATCATATGCACA | 121213 |
| Qy | 2092 | agttttacctcagccagttgttctcagagcaggcaactttttgaaactcctgctctgtg | 2151 |
| Db | 121212 | AGTTTTTACCTTAGCCAGTGTGTTCTGAGGACACAGGCAACTTTTGAACCTCCTGCTCTGTGG | 121153 |
| Qy | 2152 | agaatgtatactttatgctctctcgtgaatgtatgacaccccgcttttaaaacaaacaaacaa | 2211 |
| Db | 121152 | AGAAFTGATATCTTTATGCTCTCCGAAATGATGACACCCAGCTTTAAANA- - - - - | 121100 |
| Qy | 2212 | acaaacaaaaaagggtggggggagggaagaaagagagcctcgtcactcccttggtt | 2271 |
| Db | 121101 | --AAACAAAAAGAGGGTGGGGAGGGAGGAAAGAGAGAGAGCTCTGCACCTTCCCTTTGTT | 121044 |
| Qy | 2272 | gtagctcacagtaatacacagatatcttaattcttctaatttcccccaataatgccagaa | 2331 |
| Db | 121043 | GTAGTCTCACAGTATAACAATAATTCTAA- - -TTCCTTAATATTCGCCCAATAATCCAGAA | 120987 |
| Qy | 2332 | attggcttaatgtagcttttcactaaattctaatcctcaaaLagattgctccttaaatccaatgtt | 2391 |
| Db | 120986 | ATTGSCCTACTGATGCTTTCACTAAAT- --TCAAGTAGATTGCTCCTCAATTTCAATTGTT | 120930 |
| Qy | 2392 | aaaattggatacagaataattatcacagggaaacttaattgttaagccattagcatagaaaaa | 2451 |
| Db | 120929 | AAAAATTGGATCAGAAATAATATACAGAAGACTTTAAATGTTAAAGTCATTAGCATAGAAAAA | 120870 |
| Qy | 2452 | ctgtctcagttttatttttacctaactaacatagtaaacctaaggaagtgcgaat | 2511 |
| Db | 120869 | CTGTTCTCAGCTTTATTTTACCCTAACACGACATGAGTAACCTAAGGGAAGTGCCTGAAT | 120810 |
| Qy | 2512 | gggtgtggcaggggtattaaacgtgcatttttactcaactacctcaggtatttcagtaata | 2571 |
| Db | 120809 | GGTCTTGGCAGGTGATTAACATGCATTTTATTCACCTACCTCAGGTATTTCAGTAATA | 120750 |
| Qy | 2572 | caatgaagaacaaattgttcccttt- - - - - | 2596 |
| Db | 120749 | CAATGAAAGACAAATTTGTCCCTATTGGCGAAATTCATCCCGAATATTTTCATGTAGGTT | 120690 |
| Qy | 2597 | - - - - - | 2596 |
| Db | 120689 | GTTTTTCATTTTCCCTAAAGTGTGAGCTGTGTCAGAGAAATAAAGGGACAGAGTACAAAAAGA | 120630 |
| Qy | 2597 | - - - - - | 2596 |
| Db | 120629 | GAGAAATTTTAAAGCTAGGTGCCCGGGGAGACATCACATGTACAGCAGTTCTGTGATGC | 120570 |
| Qy | 2597 | - - - - - | 2596 |
| Db | 120569 | CCCGAGCTGTAAACCAACAATTTTTTATTAGTGAATTTTCAAAAGGGGAGGAGGTGATG | 120510 |

| | | | |
|----|--------|---|--------|
| Qy | 2597 | ----- | 2596 |
| Db | 120509 | AATAGGGTCTGGGTACACAGATCACATGCTTCAAGGTAATAAGATATACAAAGTAA | 120450 |
| Qy | 2597 | ----- | 2596 |
| Db | 120449 | ATGGAGCAGGCGGAGATCACAGGACCACAGGACTGGGGCAAAATTAAAAATTGCTAATGA | 120390 |
| Qy | 2597 | ----- | 2596 |
| Db | 120389 | AGTTTCAGCAGGCCATTGTCATTGATAACATCTTATCAGGAACAGGGTTTGAGACAGA | 120330 |
| Qy | 2597 | ----- | 2596 |
| Db | 120329 | CAACCGGTCGACCACAAAATTTATTAGCGGGAGTTCCTCGTCTCTAATAAGCCTGGGAAT | 120270 |
| Qy | 2597 | ----- | 2596 |
| Db | 120269 | GCTACGGGAGACTGGGGCTTATTTTCATCCCTACAGCTACGACCTAAAGACACGCGCCT | 120210 |
| Qy | 2597 | ----- | 2596 |
| Db | 120209 | CCAAAGCCCTCAGGGACACATTTCTTTCTCAGGGAATGTTCCCTTGCTGAGAAAAAGAAAT | 120150 |
| Qy | 2597 | ----- | 2596 |
| Db | 120149 | CAGTGGTATTTCCCATTCCTTTGAAAGAGAGAAATATGGCTCTGTTCACCCAGCT | 120090 |
| Qy | 2597 | ----- | 2596 |
| Db | 120089 | CACCAGCAGTCAGAGTTTAAGGTTATCTCTTGTTCCTGAAACATTGCTGTTATCCTGT | 120030 |
| Qy | 2597 | ----- | 2596 |
| Db | 120029 | TCTTTTTCAAGTGCCAGATTTTCATATTTGTTCAAACACACATGCTGTACAAACAATTT | 119970 |
| Qy | 2597 | ----- | 2596 |
| Db | 119969 | GTGCAGTTAACACAAATCATCACAGGTCCTGAGGTGACATACATCTCCTCAGCTTATGA | 119910 |
| Qy | 2597 | ----- | 2596 |
| Db | 119909 | AGATCATGAGATTAAGAGATTAAAGACATAGGAAATCACAAAGGTTATTGATTGGGGAAGT | 119850 |
| Qy | 2597 | ----- | 2596 |
| Db | 119849 | GATAAGTCTCCATGAAATCTTCAAAATTTATGTTTCAGAGATTGCAGTAAGACAGGCATA | 119790 |
| Qy | 2597 | ----- | 2596 |
| Db | 119789 | AGAAATTATAAAGTATTAATTTGCGGAACCTAATAAATGCCATGAANTCTTCACAATTT | 119730 |
| Qy | 2597 | ----- | 2596 |
| Db | 119729 | ATGTTCTTCATGGTTTTCAGCCGGTCCCTCCATCTGGGGTCCCTGACTTCGGGCAACA | 119670 |
| Qy | 2597 | ----- | 2647 |
| Db | 119669 | GTTCCTTAATTTTGAATAATTTTATATACTTTATTAATGATAGAGTCCAATCATATTTT | 119610 |
| Qy | 2648 | aaaaataaattaaaaattaaacagcaatcagctaaacagggcaaatgaatttaagatttttacttct | 2707 |
| Db | 119609 | AAAAATAAATTTAAATGTATACAG-AATCAGCTAACAGGCAAAATTAAGAAATTTTACTTCT | 119551 |
| Qy | 2708 | ggctgggtgacagtaaaagctggaataaatttaatttcagggttttttgaggcttttttgacacagt | 2767 |
| Db | 119550 | GSCGTATGACAGTAAGCTGGAATTAATTTTCAGGGTTTTTTTGAGGCTTTTGACACAGT | 119491 |
| Qy | 2768 | tattagttaaatacaattgttcaaaaaatacggagcagtgacctagtaactgtgagagcagac | 2827 |
| Db | 119490 | TATTA-AAAAATCAAAATGTTCAAAAGATATGGAGCAGTGCCTAATATCTCGAGAGACGAA | 119432 |


```
Qy 2828 taccattattcttcttatttagttggaaagtttttgacgggtactacaagaagtggctg 2887
Db 119431 TACCATTATTCTTTTCATTTATAGTTGAAAACGTTGTTGATGGTACTAACAAAGTGGTG 119372
Qy 2888 caggagatttggaaaggctggtttaaatggcttcaggagactcgaatttttggtttagc 2947
Db 119371 CAGGAGGCTTTGGAAACGGCTGGTTAAATGGCTTTTAGGAGACTTCAGTTTTTGGTTTACG 119312
Qy 2948 tacatgattgaatgcataataaataatgcttctgctctgactatcaatacctcaaaagaagt 3007
Db 119311 TACATGATTGAATGCATAAATAAAGCTGTGCTTTTGACCCCTTAATACCCATAGAGAAAGT 119252
Qy 3008 gcataagtgaaagagatgcaagacttcaactgactggcaaaagaacaagctttagcttgc 3067
Db 119251 GCATCAITGGAAGAGATGCAGGACTTTCAACTGATTGGCAAAAGCAAGCTTTAGCTTGTG 119192
Qy 3068 ttataggatgcttagtttggcaactacacttcagaccacaatgggaacagtcataagatgggtg 3127
Db 119191 TTACAGGTTGCTTTAGTTTGGCAGTACACTTCAGACCAATGGGACAGTCATAGATGGTGTG 119132
Qy 3128 acagtgtttaaacgcacaagaagcctacatttccatggggccagcactgtcatgagcctc 3187
Db 119131 ACAGTGTTTAAGGCACAAAAGGC-ACATCTCCATGTGGCTGGCAGCTGTCATGAGCCTC 119073
Qy 3188 actaagctatttgaagatttttaagcactgataaattaaaaaataaaataaataaattag 3247
Db 119072 ACTAAACTATTTTGAAGATTTTAAAGCAATGATAAAATTTT-----AAAAAATTAG 119022
Qy 3248 actccacttaagtag----taagataacaggatttctgtatctgtgcaatacagttc 3303
Db 119021 ACTCGCTCAAGTAGTATATAAAGTAGTAGCAGGATTTCTGTATACTGTGCAATCAGTTC 118962
Qy 3304 -tttgaaaaaaagctcaaaagatgagaatacaagaagaagttttinggataataattgaa 3362
Db 118961 TTTTITTTAAAAAAGTCAAAAGGTGAGAGAAATACAGAAAAGTTTTTGGGATATATATGAA 118902
Qy 3363 tgactgtgaaacatatgacacttggataacgaactcatttgcactcctt-----gaca 3417
Db 118901 TGACTGTGAAAAC--ATGACCTTTTGATACTGAACCTCATTTGCTCCTCTAACTGGACA 118844
Qy 3418 gcaagccagtcagtcacaatttgttgggtgtgggtgtgtgtgtgtccaaagccagctgctct 3477
Db 118843 GCAAGTCCAGTATGACAATTTGTGTGGGTGTGGGTGTGTCTCCAAGGCCACACTACTCT 118784
Qy 3478 ctgaattgatttttggatttggnttgaagatgatacacagncatgttacactgactct 3537
Db 118783 CTGAATTGACTTTGAG---TTTGTGTTGTAGATGATCAGAGTCATGTTCACAGCCTG 118727
Qy 3538 naaggacatacnttaaacacctttaaaaaaaatccccctgcctcattcttatttcgaga 3597
Db 118726 AAGG-----ATATATATCCCTTTAAAAAAATCACCTGCTCATCTCTTATTTCAAGA 118675
Qy 3598 tgaatttcgatacagactagatgtcttctt-----gaagatacaattagacattntgaa 3650
Db 118674 TGAATTTCTATACAGACTAGATATATGTTTTCTGGAAGATCATATCAATTAGACATTTTGA 118615
Qy 3651 aatgatttaaaagtgttttccctaagtctctgctgaaacaagaatttcttctttagtttaac 3710
Db 118614 AATGATTTAAATAGTTTTCCTTAATGTCTCAGAAAACAAGTTCTTTGTAGATTTTAAT 118555
Qy 3711 caaaaaagtcaccttttctgactgggtttctctcagcaattcatgatttttttttcacaca 3770
Db 118554 CAAAAAGTGCCCTTTTGTGCTACCTGGATTCACCTAGCATTCAT-AATTTTTTTTTCATACA 118496
Qy 3771 atgaattaaaaatctctaaaaatcatggactggcttcttctggttgatttcaggtagaatg 3830
Db 118495 ATGAATTTAAAAATTACTAAAAATCATGGACTGGCTTCTCGTGGTTCAGGTGAGATGTG 118436
Qy 3831 ttaaggccagagctttctcagtatgtttttttttttttccccaataattgtatttttaaaa 3890
Db 118435 TTTTAGGCCAGAGCTTTCTCAATATTTTGAATTTTTTCCCTCCCAATATTTGATTTAAAAA 118376
Qy 3891 atatacataggagctgcatttaaaaacctgctggttttaaatctctgtcanatttcacttc 3950
```

```
Db 118375 ATATACACATAGTTGCTGCATTT-ATATCTGCTGGTTTAAATTCATATATATTCACCTTA 118317
Qy 3951 tagccttttagtatggnaaatcanaatttacttacttacttaagcatttgaatttgagta 4010
Db 118316 TAGCCTTTTAGTATGCAAAATGATATTTTACTTTTACTTTAAGCATTTGTAATTTGAGTA 118257
Qy 4011 tctgtactagctaaagaataattcnaataatgagtttctactcnccaanaatgggtca 4070
Db 118256 TCTGTACTAGCTAGAAATAATTTCTATATTTGAGTTTGTACTCATCATATATGATGATCA 118197
Qy 4071 ttctcatgnataatgtntcccccaatcgacttctatcttttccagancacttgcgcaggat 4130
Db 118196 TTCTTCATCTATAATGTGCCCAATGCAGCTTCATTTTCCAGATACCTTGACGCAGGAT 118137
Qy 4131 aaatttttcatcatttagtctcccaaaaaaataaaaaaataaaaaa 4181
Db 118136 AAGGTTTTTTCATCATTTAGTGTCCCAAAAAAAGAAAGAAACTATTCAA 118086
```

RESULT 8

AC024199/c

LOCUS AC024199 188273 bp DNA HTG 12-AUG-2000

DEFINITION Homo sapiens chromosome 4 clone RP11-795024, WORKING DRAFT

SEQUENCE, 24 unordered pieces.

AC024199

VERSION AC024199.3 GI:9798101

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 188273)

Waterston,R.H.

The sequence of Homo sapiens clone

unpublished

2 (bases 1 to 188273)

Waterston,R.H.

Direct Submission

Submitted (25-FEB-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Aug 12, 2000 this sequence version replaced gi:9309539.

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: H.NH0795024

----- Summary Statistics -----

Sequencing vector: M13; 100%

Chemistry: Dye-primer ET; 99% of reads

Chemistry: Dye-terminator Big Dye; 1% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 176562 bases at least Q40

Consensus quality: 179515 bases at least Q30

Consensus quality: 180915 bases at least Q20

Insert size: 179000; agarose-fp

Insert size: 185973; sum-of-contigs

Quality coverage: 5.02 in Q20 bases; agarose-fp

Quality coverage: 5.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 24 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1759: contig of 1759 bp in length

| | | | | |
|---|---|--------|---|--------|
| D | b | 137690 | TACCATTATTATCTTTTCATTATATAGTTGAAACAGTGTGTTGATGCTACTACAAAGAGTGCTGG | 137693 |
| Q | y | 2888 | caggagattttgggaacgcgctggtttaaatggcttcaggagacttcagtttttggtttagc | 2947 |
| D | b | 137630 | CAGGAGGCTTTGGAAACGCCTGGTTTAAATGGCTTTAGGAGACTTCAGTTTGTGTTTACG | 137571 |
| Q | y | 2948 | tacatgattgaatgcataataaatgctttgctctctgactactcaatacctaagaagaagc | 3007 |
| D | b | 137570 | TACATGATTTGAATGCATTAATTAATTCGTGTGTGTTTGTACCCCTTAATACCTAGAGAAAGT | 137511 |
| Q | y | 3008 | gcacagtgaaagagatgcagagacttcaactgactgactgcaaaagcaagcttttagctgttc | 3067 |
| D | b | 137510 | GCATCATTTGAAGAGATGCAGACTTTCACTGAATTTGCAAAAAGCAAGCTTTTAGCTTTGTC | 137451 |
| Q | y | 3068 | ttaaggatgcttagtttggcaactcacactcagaccaaatgggacagctcatagatgggtg | 3127 |
| D | b | 137450 | TTACAGGTTGCTTAGTTTGGCAGTTACACTTCACACCNAATGGGACAGTCATAGATGGTGTG | 137391 |
| Q | y | 3128 | acagtgtttaaaocgcaaaaaaggtcacatttccatggggccagcactgctcatgagcttc | 3187 |
| D | b | 137390 | ACAGTGTTTAAAGGCAACAAAAGGC-ACATCTCCATGTGGCTGGCACTGTCATAGAGCCTC | 137332 |
| Q | y | 3188 | actaagctattttgaagatattttaagcactgataaattataaaaaaataaaataatag | 3247 |
| D | b | 137331 | ACTAAACTATTTTGAAGATTTTAAAGCAATGATAAATTTT-----AAAAAATTAG | 137281 |
| Q | y | 3248 | actccacttaagtag---taaaatataacaggaattctctatactgtgcaatcagcttc | 3303 |
| D | b | 137280 | ACTCTGCCCTAAAGTAGTATATAAAGTATACGAGGATTTCTGTATATCTCTGCAATCAGTTC | 137221 |
| Q | y | 3304 | -tttgaaaaaaagtcacaaagatagagaaatcacagaagaagttttingggataaaattgaa | 3362 |
| D | b | 137220 | TTTTTTTTTAAAAAAGTCAAAAAGGTAGAGAAATACAAGAAAAAGTTTTTGGGATATAATTGGAA | 137161 |
| Q | y | 3363 | tgactggaagacaatagacctttgataaagcaactcatttgcactcctt-----gaac | 3417 |
| D | b | 137160 | TGACTGTGAAAAAC--ATGACCTTTGTATCTAGTAACTCATTTGGCTCACTCTCTTAACCTGGACA | 137103 |
| Q | y | 3418 | gcaagccacagtcagtcacaattgcttgggtgtgggtggtctccaaaggccacgcgtcctc | 3477 |
| D | b | 137102 | GCAAAGTCCAGTAGTACAAATGTGTGGGTGGGTGCTCCAAAGGCCACACTACTCT | 137043 |
| Q | y | 3478 | ctgaattgatttttggagtttggnttgnnaagatgatcacagncaatggtcacactgatctt | 3537 |
| D | b | 137042 | CTGAATTTGACTTTTGAG---TTTTGTGTTTGAAGATGATCACAGTCATGTTACACTGACCTG | 136986 |
| Q | y | 3538 | naaggacatatnttataaaccctttaaaaaaaataacccctgcctcaattcttattctcgaga | 3597 |
| D | b | 136985 | AAGG-----ATATATATCCCTTTAAAAAAAATCACTCTGCCCTCACTTTATTTTCAAGA | 136934 |
| Q | y | 3598 | tgaatttcgatcacagactagatgctcttc-----gaagatcaattagacattntgaa | 3650 |
| D | b | 136933 | TGAATTTCTATACAGACTAGATATGTTTTTCTGAAAGATCATATCAATTTAGACATTTTGAA | 136874 |
| Q | y | 3651 | aatgatttaagtgctttctaatgtctctcgaaaaacaagttctcttttgtagtttttaac | 3710 |
| D | b | 136873 | AATCATTTAAAAATGTTTTTCTTAATGTCTCAGAAAAACAAGTTCTTTTGTAGTTTTAAT | 136814 |
| Q | y | 3711 | caaaaaagtgcccttttggctactggtttctctctcctagcattcatgatttttttccacaca | 3770 |
| D | b | 136813 | CAAAAAAGTGCCCTTTTGTGTGCTAGCTGGATTTCCACTAGCATTTTCAT-NAATTTTTTTCATACA | 136755 |
| Q | y | 3771 | atgaattaaagtgctaaacaatcagcctggtctctcgttggaattccaggtaaatgtg | 3830 |
| D | b | 136754 | ATGAATTTAAAAATTACTAAAAATCATGAGCTGGCTTCTCTGGTTGGATTTTCAGGTGAGATGTG | 136695 |
| Q | y | 3831 | tttaaggccacagactttcttcagtattttgattttttttcccccaataattgattttttaaaa | 3890 |
| D | b | 136634 | ATATPACATAGTTGTGTGCATTT-ATATCTGCTGGTTTAAATCTATATATATTTACATTA | 136576 |
| Q | y | 3891 | atacacacataggagctgcatttaaaacctgcgtgggtttaaattctgtcnaatttcacttc | 3950 |

| | | | |
|----|--------|--|--------|
| QY | 3951 | tagcccttttagtaggcnaatcanaaatttactcttttacttaagcatttgtaatttgagta | 4010 |
| DB | 136575 | | |
| DB | 136575 | TAGCCCTTTTAGTATATGCGCAATGATATTTTACTTTTACTTAAGCATTTGTAATTGGAGTA | 136516 |
| QY | 4011 | lctggtactagctaagaaaaaattcnaataattgattgatttgcattcctcnccaanaatgggtca | 4070 |
| DB | 136515 | TCTGGTACTAGCTAAGAAATAAATTTCTTAATTTAGTTTCTGTTTCTACTCATATATATGGATCA | 136456 |
| QY | 4071 | ttctctcgnatnaaatgtncccccaatgcagcttcattcttccaganaaccttgacgcaggat | 4130 |
| DB | 136455 | TTCTCTCATCTATAATGTGCCCCAAATCGAGCTTCATTTTCCAGATACCTTTGACGCAGGAT | 136396 |
| QY | 4131 | aaatttttctatcatttagtgcccccaaaaaaataaaaaaataaaaaa | 4181 |
| DB | 136395 | AAAGTGTTTTTCATCATTTAGGTGCCAAAAAAGAAAAAGAAAGAACTATTTCAA | 136345 |

| | |
|--------------|---|
| RESULT | 9 |
| LOCUS | AF042353 |
| DEFINITION | Xenopus laevis KH domain-containing transcription factor B3 mRNA, 22-JAN-1998 |
| ACCESSION | AF042353 |
| VERSION | AF042353.1 |
| KEYWORDS | GI:2801765 |
| SOURCE | African clawed frog. |
| ORGANISM | Xenopus laevis |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Pipidae; Xenopodinae; Xenopus. |
| AUTHORS | Pfaff,S.L. and Taylor,W.L. |
| TITLE | Characterization of a Xenopus oocyte factor that binds to a developmentally regulated cis-element in the TFIIA gene |
| JOURNAL | Dev. Biol. 151 (1), 306-316 (1992) |
| REFERENCE | 1 (bases 1 to 3147) |
| MEDLINE | 92249652 |
| AUTHORS | Griffin,D. and Taylor,W.L. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (11-JAN-1998) Biochemistry, University of Tennessee, Memphis, 858 Madison Ave., Memphis, TN 38163, USA |
| FEATURES | Location/Qualifiers |
| source | 1..3147 |
| | /organism="Xenopus laevis" |
| | /db_xref="taxon:8355" |
| | /cell_type="immature oocyte" |
| | 972..2753 |
| CDS | /codon_start=1 |
| | /product="KH domain-containing transcription factor B3" |
| | /protein_id="AA897457.1" |
| | /db_xref="GI:2801766" |
| | /translation="MNKLXIGLNSENVSPDLESIFKSKIPFTGQFLVKSGYAFV . PDETWMKAIDTSLGKVELHGVIEVHSPKRRSRKLIQRNIPPHLQWEVLDSLQ . QYGVTEQOVNTDSETAVVNTYANKHAROGLEKNGYOLENTSLKVTYIPDEM . PQSPQLOOQPOOHQPGRRGSGPARQSGPAAARPKQSEVPLRMLVPTQFV . IIGKEGATRNITKQTSKIDIRHENAGAEEKPTIHTPECCSAACKIIMEINQ . AQDTKEETPILKAHNPGVLRGKGRNKKLEQDTRKITSIPLODITLYNIP . TITVKSISTCAKAEVEVMKKITREYENDIAAMNLOAHILPGLNLALGLDLPSSSS . PPPSAGVSPPTTSASYPFGQOPESETVHLFPALAVGAIIGKQGHQIKQLSRFAG . IKTAPAGDPDAKURMYITTPGPEAQKQAGRIYGLKLEENFGPEEYKLEAKIKV . YAAGRVIGKGGTIVNELQNTSAEVVVPDQTPDENDQVVKITGHEFVASQLAQRK . EILAQVRQOQOQOQKTAQSGQPQPRK" |
| misc_feature | 1607..1723 |
| | /note="encodes KH domain 1" |
| misc_feature | 1847..2002 |
| | /note="encodes KH domain 2" |
| misc_feature | 2243..2389 |
| | /note="encodes KH domain 3" |
| misc_feature | 2489..2641 |
| | /note="encodes KH domain 4" |
| BASE COUNT | 957 a 690 c 676 g 824 t |
| ORIGIN | |

| Query Match | | 26.7%; | Score 1117; | DB 5; | Length 3147; |
|----------------------------|------|--|---------------------|--------|--------------|
| Best Local Similarity | | 73.7%; | Pred. No. 4.1e-215; | | |
| Matches 1622; Conservative | | 0; | Mismatches 510; | Indels | 69; Gaps |
| Qy | 211 | tttttctcgttattttttttaaacacaccactcttcacaatgaacaaactgtatatcgg | 270 | | |
| Db | 932 | TCITTTTAGCCTTTTATTTGGGTTTTTTTTTTATTTTTCAGATGAACAAGCTGTATATGG | 991 | | |
| Qy | 271 | aaacctcagcgagaaacgcgcgccttcgacactgaagaagtattcttcaaggagcgccaagat | 330 | | |
| Db | 992 | AAACCTAAGCGAAACGTCAGCCCCCGGATTTGGAAAGTCTCTTCAAGGAGTCAAGAT | 1051 | | |
| Qy | 331 | cccggtgtcgggacccttctcgtgtgaagactggctacgcgttcgtggactcccgagacga | 390 | | |
| Db | 1052 | CCCTTTCACCGGCAGTTTCTGGTCAAAAAGTGGATACGCGTTTGTGGACTGTCCGGATGA | 1111 | | |
| Qy | 391 | gagctgggcccctcaaggccatcagggcggtttcaggttaaaatagaactgcacgggaacc | 450 | | |
| Db | 1112 | GAGCTGGGCTTATGAAGGCCATCGACACCCCTCTCAGGGAAGTGGAGTTCATGTCACAAAGT | 1171 | | |
| Qy | 451 | catagaagttgacactcgggtcccaaaaggcaaaaggattcggaaaacttcagatcacgaaa | 510 | | |
| Db | 1172 | GATAGAAGTTGAACATCTAGTACCCAAAAGGCAAGAGGTGGAAGCTTTCAGATAAGGAA | 1231 | | |
| Qy | 511 | tatccgcctcatttaccgtggagagtgctgatatagtttaactagttccagttagagtggt | 570 | | |
| Db | 1232 | CATACCACCTCATCTACAGTGGGAGTACTGGACAGCCCTATTAGCACAGTATGGGACAGT | 1291 | | |
| Qy | 571 | ggagagctgtgagcaagtgaacactgactcggaaaactcgacttgttaaatgtaacctattc | 630 | | |
| Db | 1292 | GGAAAACGTGTAGCAAGTTTAACACTGATTCAGAAACTGTCAGTGTAGTAATGTAAACATATGC | 1351 | | |
| Qy | 631 | cagtaagaacccaagcttagacaagcactagaacaaactgaatggatttcagttagagaattt | 690 | | |
| Db | 1352 | CAATAAGGAGCACGCTTAGCAAGAGCTCGAAAATTTAAACGGCTATATCAGCTGTGAACATA | 1411 | | |
| Qy | 691 | caccttgaagtagccttatatccctgatgaatggccgccagcaaaaaccccttgacga | 750 | | |
| Db | 1412 | TAGCCTGAAGAGTCAGCTATATACCTGTATGAATGGCAACACCGCAANTCACCATCCGACGA | 1471 | | |
| Qy | 751 | g-----cccccagagtgccgcggggggttgggcgagaggggtc | 786 | | |
| Db | 1472 | GCTTCAGCAGCCGACGACGACCAACACCTTCAGGGGAGGCGTGGGTTTGGCAAGGGGACC | 1531 | | |
| Qy | 787 | ctcaaggcagggtgtctcaggatccgtatcccaagcagaaaaac---atgtgatttgctctc | 843 | | |
| Db | 1532 | TGCCAGGCAGGGTCCCGCTGGTGCTGCTGCACAGACCTTAAGCCTCAGTCAGAGGTTCACCT | 1591 | | |
| Qy | 844 | gcgcctcgtggttcccaaccctaattgttgagccatcatagaaaagagatgccacat | 903 | | |
| Db | 1592 | GAGGATGCTGGTCCCAACACAGTTTGTGGGCGGATCATTTGAAAAGAGGGGTGCCACAT | 1651 | | |
| Qy | 904 | tcggaacatcaccaaaacagaccagtcctaaaaatcgatgtccaccgttaagaaaaatgcggg | 963 | | |
| Db | 1652 | CAGGAACATCACTAAACAGACACCAGTCAAAAATAGATATTCAACCGTAAAGAAATGCTGG | 1711 | | |
| Qy | 964 | ggctgtgagaadtgatattactatctctactcctctgaaggcaccctcgcgggttgtaa | 1023 | | |
| Db | 1712 | TGCTGCAGAGAAAACCGATTACGATTTCACTCAACCCCTGAAAGCTTTCAGCTGCATGCGAA | 1771 | | |
| Qy | 1024 | gtctattctgagatttatcataaggaagctcaagatatataaaattcacagaagatccc | 1083 | | |
| Db | 1772 | GATCATTATGGAGATCATGCGAAGGAAGGACTCAGGATACCAAGTTTCATGAAGAAATCCC | 1831 | | |
| Qy | 1084 | cttgaagattttagctcataaataactttgttgagcgtcttatttggttaagaaaggaagaaa | 1143 | | |
| Db | 1832 | CTTAAAAAATATTAGCACATAACAAATTTTGTGTGGCGCTCTATTGGGAAGAAGGAGGAA | 1891 | | |
| Qy | 1144 | tcttaaaaaaattgagcaagacacagacactaaaaatcacgatatactctcattgcgggaatt | 1203 | | |
| Db | 1892 | CCTTAAGAAAATAGACAGACACAGATACCAAAATACAAATATCTCCACTTACAGGACTT | 1951 | | |

| | | | |
|----|------|--|------|
| Qy | 1204 | gacgctgataaatccagaagcgaactattacagttaaaaggcaatggttgagacatgtgcaaa | 1263 |
| Db | 1952 | GACACTGTACAACTCTGAACCGGACCAATTACAGTAAAGAGCGACATGTGAGCATGTGCGAA | 2011 |
| Qy | 1264 | agctgaagagagatcatgaagaaatcaggagctctatgaaatgatattgcttctat | 1323 |
| Db | 2012 | AGCCGAGGAAGAGTCATGAAAAAATTAGGAGGTCCCTATGAAATGATATTGCTGCTAT | 2071 |
| Qy | 1324 | gaattctcaagcacatttaattctctgga ttaattctgaacgcttggtgtctgtttccacac- | 1382 |
| Db | 2072 | GAATCTACAGCACATTGTATTCTGGTTTAAACCTGAATGCATTGGGCCCTTTTCCCAAC | 2131 |
| Qy | 1383 | --ccattcagggatgccacctccacctc- --agggcccocttcagcca tgaactctcc | 1437 |
| Db | 2132 | ATCCCTGTCAGGAATGCCACCACTCTGCTGGAGTTCTCTCGGACAACATCTGCTTC | 2191 |
| Qy | 1438 | ctaccgcagtttg---agcaatcagaagaacgagactgttctcaagtttatcccagctct | 1494 |
| Db | 2192 | TTATCCACCATTTTGGGCAGCGACGACAGTTCAGAGACTGTTCATCTTCCTATCCCAAGTTT | 2251 |
| Qy | 1495 | atcagtcggttgccatcatcggaacgagcgccagcacatcaagcagcttctcgccttgc | 1554 |
| Db | 2252 | AGCTGTGGGAGCAATTTATTGGAAAGCAAGACAACACATCAAAACAACCTGTCGCTTTTGC | 2311 |
| Qy | 1555 | tggagcttcaattaaagattgctccagcggaagcaccagatgctaaagtggagatgggtgat | 1614 |
| Db | 2312 | AGGAGCATCTATTAAAGATTGCACCCGACAGGGACCTTGATGCCAAACTAAGAATGGTGAT | 2371 |
| Qy | 1615 | tatcctggaaccaccagaggtcagttcaaggtctcagggaagaatttatggaaaaataa | 1674 |
| Db | 2372 | CATCACTGGACCTCTCTGAAGCCCAATTTAAGGCTCAAGGAAGGATCTATGGTAAACTTAA | 2431 |
| Qy | 1675 | agaagaaacttgttagtctctaaagaagagtgaaacttgaagctcatatcagagtgcc | 1734 |
| Db | 2432 | AGAAAGAAACTCTCTTGGACCTTAAGAAAGAACTGAAACTTTAGGCCCCACATTAAGATGCC | 2491 |
| Qy | 1735 | atcctttgctgtcgtggcagagttatttggaaaaggagcaaacggtgaaatgaacttcagaa | 1794 |
| Db | 2492 | GTCATATGCTGTGGACGTGTATTGGCAAGAGGAGGCAAAACAGTAAATGAAC TTCAGAA | 2551 |
| Qy | 1795 | tttgcagtgcagaagtgtgttcctcgtgaccagacacctgtagaatagccaaagt | 1854 |
| Db | 2552 | TTTAAACAAGTGCAGAAGTTGTTGTGCCCGTGTATCAAAACCCAGATGAAATGATCAAGT | 2611 |
| Qy | 1855 | ggttgctcaaaataacttggctcaacttctatgcttgcaggttgcccaagagaaaaattcagga | 1914 |
| Db | 2612 | GGTTGTTTAAGATAACCGGTCTACTTCATGCTTAGCCAGCTTGCCACAAAGGAAAAATCAGGA | 2671 |
| Qy | 1915 | aattctgactcaggttaa-----agcagcaccaacaacagaggtctctgcaagtggacc | 1968 |
| Db | 2672 | AATACTGGCTCAGGTATAGAGACAGCAGCAGCAACAGCAGAGACAGCGCAAGATGGACA | 2731 |
| Qy | 1969 | acctcagtcgaagcgaagttaaaggctcagg-aaacagccaccacacagagcagatgcc | 2027 |
| Db | 2732 | ACCCCAACCAAGAAGAAAAATAAGGTTTCAGAAAAATGGCCCTTTCACAGACAGATGCCA | 2791 |
| Qy | 2028 | aaccaagacagatgtccttaaccaacagatggggcgtgaccccttatccagaatacaatg | 2087 |
| Db | 2792 | GACCAAAAGACAGACTGAGATTAAGACAGAGTCTCCACTCTGTGTGCTACGTGGGAGGAA | 2851 |
| Qy | 2088 | cacaagttttactcagcaggtgttcttgaggaaccagcgaactttgaactctgtctc | 2147 |
| Db | 2852 | AACCACCTAGTACATCCCTTGTT-----TGAGGACCAAGCAACCAATTCGAAC-----TGTC | 2901 |
| Qy | 2148 | tgtgagaaagtatacttta tgcctctgtaaatgtatgacaccacagctttaaacacaa | 2207 |
| Db | 2902 | TCTGAGAATGTATTATTAGGCTCTC-AAAAC TTTTGAACCCAGCCTTTAA---AAATGG | 2957 |
| Qy | 2208 | acaaacaaacaaaaaagggtgggggagggaggggaagagagctctgcacttccctt | 2267 |
| Db | 2958 | AAAAAAAAGGCAAAAAAGAAATGAAAAAATAAAAAAATAAAGGTCCGCTTCACTTCCCTTT | 3017 |

| | | | |
|------------|----------|---|-----------------|
| Db | 1459 | ATCCTCGTCAGGATGCGACACACATCTGCTGGAGTTTCCTCTCCGACACACATCTGCTTC | 1515 |
| Qy | 1438 | ctaccgcgagttgt---agcaatcagaaacagagactgttcatcagtttatcccgactct | 1494 |
| Db | 1519 | TTATCCACCATTTGGGCAGCAGCCAGATGTCAGAGACTGTTCATCTCTCATCCGAGCTT | 1578 |
| Qy | 1495 | atcagtctggtgcatactatcggcgaacagggccagcaatcaagcagctttctcgtcttgc | 1554 |
| Db | 1579 | AGCTGTGGGAGCAATTTATTTGAAAGCAAGGACACACATCAAAACAACATGTGCGGCTTTGC | 1638 |
| Qy | 1555 | tggagcttcaataaagattgctccagcggaagaccagactgtctaaagtggagatggtgat | 1614 |
| Db | 1639 | AGAGCATCTATTAGATTTTGCACCCGACAGAGGACCTGATGCCAAACTAGAAATGGTGAT | 1698 |
| Qy | 1615 | tatcactggaccaccagagctcagttcgaagctcaggaagaatttatggaataataa | 1674 |
| Db | 1699 | CATCACTGGACCTCCTGAAGCCCAATTTAAGGCTCAAGGAGGATCTATGTTAACTTAA | 1758 |
| Qy | 1675 | agaagaaactttgttagtccttaagaagaggtgaaacttgaagctcatatcagatgccc | 1734 |
| Db | 1759 | AGAAGAAAACTTCTTTGGACCTTAAAGAAAGATGAAACTTCAGGCCACATTTAAAGTGCC | 1818 |
| Qy | 1735 | atcctttgctgctgcagaagttattgaaaaggagcaaaacgtgaatgaacttcagaa | 1794 |
| Db | 1819 | GTCAATATGCTGCTGGACGTGTTATTGGCAAGAGGCAACAGTAATACTGAACCTTCAGAA | 1878 |
| Qy | 1795 | tttgtcaagtgcagaagttgttgcctcgtgaccagacacctgatgagaatgaccaagt | 1854 |
| Db | 1879 | TTTAAACAAGTCAGAGTTGTTGTGCCCGGTGATCAACACCCAGATGAAATGATCAAGT | 1938 |
| Qy | 1855 | ggttgcaaaaataactggctactcttatgcttgcaggtgtccagagagaaaaattcagga | 1914 |
| Db | 1939 | GGTTGTTAAGATAACGGGTCACTTCTATGCTAGCCAGCTTGCACAAAGGAAAAATTCAGGA | 1998 |
| Qy | 1915 | aattctgactcaggtaa-----agcagcaccacaacagaaagctctgcaaaagtggacc | 1968 |
| Db | 1999 | AATCACTGGCTCAGGTAAGAGACAGCAGCACAGCAGAAAGACACGCAAAAGTGGACA | 2058 |
| Qy | 1969 | acctcagtcaagcaggaagttaaaggctcagg-aaacagccccaccacagagcagatgccca | 2027 |
| Db | 2059 | ACCCCAACCAAGNAGAAATAAAGGTTCAGGNAATGGCCCTTCACAGACAGATGCCCA | 2118 |
| Qy | 2028 | aaccaagacagattgctttaaccaacagatggcgctgacccccctatccagaatcacatg | 2087 |
| Db | 2119 | GACCAAAAGACAGACTGAGATAAAGACAGGTTCCACCTCTGTCTGTACGTGGGAGGAA | 2178 |
| Qy | 2088 | cacaagttttaccttagccagttgtttctgaggaccaggaacattttgaaactctctctc | 2147 |
| Db | 2179 | AACCACTAGTATCATCCCCCTGTT----TGAGGACCGACGGCAACCATTTGAAC-----TGTC | 2228 |
| Qy | 2148 | tgtgagaatgtataactttatgctctctgaaatgtatgacccccagcttttaaacacaaaca | 2207 |
| Db | 2229 | TCGTGAGAAATGATTATTAGGCTCTC-AAAACCTTTTGAAACCCAGCTTTAA--AAATGG | 2284 |
| Qy | 2208 | acaaacaaacaaaaaggggtgggggaggggagagagagagagctctgcacttccctt | 2267 |
| Db | 2285 | AAAAAAAAGGCAAAAAAGAAAAATGAAAAAATAAGGTCCCTTCATCTCCCTTT | 2344 |
| Qy | 2268 | tgtgtagctcacagataaacagatatcttaattctcttaattatcccccaataatgccc | 2327 |
| Db | 2345 | TGATCTATTCTCAAAAGTTAAACATAAACCTGTGTAGTCTTTTTTTTATTGTCTTATTCC | 2404 |
| Qy | 2328 | agaaattggcttaatgatgctttcactaaa | 2357 |
| Db | 2405 | CCCAACAATTTAGAATCTTGTTTAAATGAA | 2434 |
| RESULT | 11 | | |
| LOCUS | AP002376 | | |
| DEFINITION | AP002376 | 161404 bp DNA | PRI 21-JUN-2001 |
| ACCESSION | AP002376 | Homo sapiens genomic DNA, chromosome 11q, clone: RP11-349I116, complete sequence. | |

```

VERSION      AP002376.3  GI:14517592
KEYWORDS
SOURCE      Homo sapiens DNA, clone:RP11-349I16.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens genomic DNA
JOURNAL
REFERENCE   2 (bases 1 to 161404)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL
COMMENT     Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            On Jun 20, 2001 this sequence version replaced gi:11094148.
FEATURES
            Location/Qualifiers
                source
                1..161404
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="11"
                /map="11q"
                /clone="RP11-349I16"
BASE COUNT  47396 a 33453 c 33200 g 47355 t
ORIGIN
Query Match      25.9%; Score 1084.4; DB 9; Length 161404;
Best Local Similarity 88.4%; Pred. No. 1.4e-208;
Matches 1305; Conservative 0; Mismatches 127; Indels 44; Gaps 11;

Qy 2730 aaataatttcagggttttttgagcgttttgacacagttattagtt---aaatcaaatgt 2786
Db 130573 AAATTAATTTTCAGGGTTTTTTTTCAGGCTTTTCACACAGTTATTAGCTAAAAAATCAAAATGT 130632

Qy 2787 tcaaaaatacaggacagtcgctagtatctggagagcagcactaccatttattcttcatt 2846
Db 130633 TCAAAGATATATGGACAGTGCCTAATGTCTGGAGAGCAGCATTACCATTTATTATTTCATT 130692

Qy 2847 tatagttgggaaaagtttttcgagctactaa-----caaaagtggtcgaggagattttgga 2901
Db 130693 TATAGTTGGGAAAAGTTTTTGATGTACTAACAGATCAAAGTGGTGGCAGAGGTTTTCGA 130752

Qy 2902 acggtgggtttaaatggcttcaggagacgttcagttttttgttagctacatgattgaatg 2961
Db 130753 ACGGCTGGTTTTAAATGGCTTCAGGAGACTTCAGTGTGTTTTGTTTAGCTACATGATTGAATG 130812

Qy 2962 cataataaattgctttgtctctgactatcatacctaaagaagaagtcagtgaaagag 3021
Db 130813 CGTAATAAATGCTTTGTGCTTTTGACTATCAATGCCTAGAGAAAAGTGCATCAGTGAAGAG 130872

Qy 3022 atgcaagaacttcaactgactggcaaaagcaagcttttagctgtgcttataggatgctta 3081
Db 130873 ATGCAGGACGTTCAACTGATTTGGCCAAAAGCAAGCTTTTAGCTTGTCTTACAGGATGCTTA 130932

Qy 3082 gtttgccactacacttcagacacaattgggacagtcatagatggtgtgacagttgtttaaacg 3141
Db 130933 GTTTGCCAGTACACTTCAGACCAATGGGATAGTCATAGATGCGGTGACAGTGTTTTAAAGG 130992

Qy 3142 caacaaaggctacatttccatggggccagcactgtcatgagcctcactaagctattttg 3201
Db 130993 CAACAAAAGGCTACATCTCCATGTGTGGCCAGCAGCTGTCTATGAGCCCTCACTAAGCTATTTTG 131052

Qy 3202 aagatttttaagcactgaataattaaaaaataaaataatagactccaccttaagt 3261
Db 131053 AAGATTTTTTAGCAATGATAAATTA-----AAAAAATATATATCCCCACCTTAAAGT 131103

Qy 3262 ag----taaaactaacaacgaattttctgtatactgtgcaatcagttcttttgaaaaaaagt 3317

```


Db 131104 AGTACATAAAGTAGACAGGATTTCCGTATACTCTGCAATCAGTCTCTTGAAGAAAAGT 131163
Qy 3318 caaagaatagagaatacaagaagaatttngggataataattggaatgactgtgaaacat 3377
Db 131164 CAAAAGTAGAGATAACAAGAAA--TTTTTGGGATATAATTTGAATGACTGTGAAAC-- 131219
Qy 3378 atgacatttgataacgaactcatttgtcactc-----cttgacagcaaacccagtaag 3432
Db 131220 ATGACCTTTGATCTGACATCATTTGTCTACTCTTAACTTGACAGCAAGGCCAGTATG 131279
Qy 3433 tacaattgtgtgggtgtgggtgtctcccaaggccagctgctctcgaattgatttttt 3492
Db 131280 TACAACGTGTGACTGTGGGTGCTCTCAAGGCCAGCTGCTCTCTGAATGATTTGA 131339
Qy 3493 gaggtttggntgnaagatgatcacagncatgtttacactgatctttnaaggacatatnta 3552
Db 131340 G---TTTTTGTGTGAAGATGATCAGATCATGTTTACACTGACCTGAAGGCATATATATA 131396
Qy 3553 taaccttttaaaaaaaatccccgcctcattcttatttcgagatgaattcgatacag 3612
Db 131397 CATAACACTTTTAAAAAAACACTCTGCCTCATCTTTTCAAGATGAATTTCTATACGG 131456
Qy 3613 actagatgtcttctc-----gaagatcaattagacattntgaaatgatttaagt 3663
Db 131457 ATTAGATATGTTTTTCTGAAGATCATATCAATTAGACATTTTGAATAATGACTTAAAT 131516
Qy 3664 gtttccctaatgtctctctgaaacaaagtcttcttttagtttttaacccaaaaaagtcccc 3723
Db 131517 GTTTTCTTAATGTTCTCAGAAAAAAGTTCTTTTCTAGTTTAAACCAAAAAGTGCC 131576
Qy 3724 tttttgcaactggttctctctagcattcagtgatttttttttttttttttttttttttttt 3782
Db 131577 TTTTGTGCTGGATTCACCTAGCATTCATAAATTTTTTTTTTTCATACATGAATTTAAAT 131636
Qy 3783 tgcataaatcatgactggtcttctggtgagttcaggttaaatgttttaagggcaga 3842
Db 131637 TGCTAAATCATGACATGGCTTCTGTGTTGATTTCAAGGTGAGATGTGTTTAAGGCCAGA 131696
Qy 3843 gctttctcagttattgattttttttccccaatattttgatttttttaaaatatatacacatag 3902
Db 131697 GCTTTTCTCAATATTTGATTTTTTTCTCAATATTTGATTTTTTAAAAATATACACATAG 131756
Qy 3903 gactgcatttaaaacccgtggttttaaatctctgcanattcactctcgtccttttagt 3962
Db 131757 GTGCTGCAATTT-ATATCTGCTGTTTAAATCTGTCTATATTTTCACTTCTAGACTTTTAGT 131815
Qy 3963 atgcnnaatcanaatttacttttacttaagcatttgttaatttgagatctgtgactagc 4022
Db 131816 ATGCCAATCATATTTTACTTTTACTTTAAGCATTTGTAGTTTGGAGTATCTGGTACTAGC 131875
Qy 4023 taagaaataattcnaataattgagtttgttactcncnaaanatgggtcattctcctcatgnat 4082
Db 131876 TAAGAAATAATCTATAGTTGAGTTTCTGCTCACCATATATGGATCATCTTCTCATCTAT 131935
Qy 4083 aatgtncccccatgcagcttcattttccagahanaccttgacgaggataaatttttcat 4142
Db 131936 AATGTGCCCAATATGACGTTCATTTTCCAGATACCTTGACGCAGAGTGAAGTTTTTCAT 131995
Qy 4143 catttaggtcccaaaaaaataaaaaaataaaaaa 4178
Db 131996 CTTTGTAGTGCACAAAAAATAAAAAAAGGGCAAGA 132031

RESULT 12
AF064634
LOCUS AF064634 2405 bp mRNA VRT 03-JUN-1998
DEFINITION Xenopus laevis vgl RNA binding protein variant D mRNA, complete cds.
ACCESSION AF064634
VERSION AF064634.1 GI:3172448
KEYWORDS African clawed frog.

ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 2405)
AUTHORS Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman Schwartz,S., Standart,N.M. and Yisraeli,J.K.
TITLE RNA binding protein conserved in both microtubule and microfilament-based RNA localization
JOURNAL Genes Dev. (1998) In press
REFERENCE 2 (bases 1 to 2405)
AUTHORS Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman Schwartz,S., Standart,N.M. and Yisraeli,J.K.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) Biochemistry, Cambridge University, Tennis Court Road, Cambridge CB2 1GA, UK
FEATURES
source 1..2405
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/dev_stage="oocyte"
379..2163
/note="Vgl-RBP; Vgl localization element binding protein"
/codon_start=1
/product="Vgl RNA binding protein variant D"
/protein_id="AACI8598.1"
/db_xref="GI:3172449"
/translation="MNKLYGNLSENVSPDLSLFKESKIPFTGQFLVKSQYAFVDC PDETWAKAIDTSLGVELHGVIEVESHVPKRQRKQIRNPFIHLOWEVLDSLLA OYGTVCNQVNTSEAVNVNTYANKEHAROGLEKINGVOLENSILKVTYIDENAT PQAPSQLOQPOQHPQRRGFGQGRGAPGSGAARPKPQTEVPLRLMLVPTQFVG AIIKGGATIRNITKQSKIDIRKENAGAEEKPIITHSPGCSAAACKILMEIMOK EAQDTFTPEIKLILAHNNFVGLJKEGNLKKIEQDTHKTITSLQDLFLYNPE RTTIVAGSIEPCAKAEEEMKIKRESYENDIAAMNQALHPIGLINLALPFPSSSG MPSPSVGSPSTSTSPFPQOPESETVHLFIPALAVGAIIGQGOHIKQLSRFAGA STIKIAEPGDAKLRLVITGPPEAFKAGRIYGLKKEENFGPEEYKLETHIKVP SYAAGRVIGKGTYNELNLTSAEVVVPDQTPDENDEVVKITGHFYSQASLAQRKI QETLAQVRKQQQQQKTVSGQPQRRK"
BASE COUNT 763 a 537 g 537 c 568 t
ORIGIN
Query Match 25.9%; Score 1084.2; DB 5; Length 2405;
Best Local Similarity 73.2%; Pred. No. 1.7e-208;
Matches 1521; Conservative 0; Mismatches 503; Indels 55; Gaps 8;
Qy 192 ggtgggttaatttcgtttttctcgtttatttttttaaaacaccactcttcacaa 251
Db 320 GGCTCGCTCTCTTTAGCTTTTATTATTATTTTATTTTATTTTATTTTACAA 379
Qy 252 tgaacaaactgtatctcgaaacctcagcagagaacgcgcgccctcgaccagaaagta 311
Db 380 TGAACAAGCTGTATATTGGAACCTCAGCGAAAAACGTGAGCCCCACGGATTGGAAGTC 439
Qy 312 tcttaaggacgcgaagatccccgggtgtcgggacccttccctggtaagactggctacgcgt 371
Db 440 TCTCAAGGAGTCTAAGATCCCTTTCACCGGCGAGTTTCTGGTCAAAAGTGATACCGT 499
Qy 372 tctgtgactcccgacgagactgggcccctcaagccatcagcgccgtcttcaggtaaaa 431
Db 500 TTGTGGACTGTCCGATGAGACGTGGGCTATGAAGGCCATCGACACCTCTCAGGGGAAG 559
Qy 432 tagaactgcacgggaacccatagaagttagcactcgggtcccaaaaaggcaaggattc 491
Db 560 TGGAGTTGCATGGCAAAAGTGATTGAAGTCGAACACTCGGTACCCAAAAAGGCAAGGAGTC 619
Qy 492 ggaacttcagatacagaaatattcccgctcatttacaagtgggaggtgctgagtagttac 551
Db 620 GAAAGCTTCAGATACGAAACATACCCGCTCATCTACAGTGGAGGTACTGACAGCCGTGT 679
Qy 552 tagtccagtagtgagtggtgagagctgtgacaaagtgaacactgactcgggaactcgag 611
Db 680 TGGCACAGTATGGCAGAGTGGAAAACTGTGCGCAAGTGAACACTGAATCAGAACTGCAG 739

| | | | |
|----|------|--|------|
| Qy | 612 | ttgtaaatgtaacctattccagtaaggaccgaagctagacaagagcactagacaaactgaatg | 671 |
| Db | 740 | TAGTGAATGTAAACATATGCCAATAAGGAGCAGCGCTAGACAAGGACTCGAAAAATTAAACG | 799 |
| Qy | 672 | gattcctgttagagaattccacttgaaagtgcctctatccctgtagtgaatggccgcc | 731 |
| Db | 800 | GCTATCAGCTGGAAAACATATACCTGAAAGTGCACGTATATACCTGTGTGTAATGGCAACAC | 859 |
| Qy | 732 | agcaaaacccct-----tgcacgacgcccgagggtcgcc | 764 |
| Db | 860 | CGCAAGCACCATCCCAACAGCTTCAGCAGCAGCGCAGCAACACCCCTCAAGGGAGGC | 919 |
| Qy | 765 | gggggcttgggcagagggctcctcaaggcagggtctccaaggatcogtattccaagcaga | 824 |
| Db | 920 | GTGGGTTTGGGCAAGGGGACCTGCCAGCGAGGGATCCCTGGTGCCTGCTGCAAGACCTA | 979 |
| Qy | 825 | aacc--atgtatttgctctgcctgcctggttcccccacccaatttgttggagccatca | 881 |
| Db | 980 | AGCCTCAGACAGAGGTTCCGCTGAGAACTGCTGGTTCCCAACACAGTTTGTAGGGCGGATCA | 1039 |
| Qy | 882 | taggaaaaagaggtgcacacatctcggaacatccacaaacagaccagctctaaaatogatg | 941 |
| Db | 1040 | TTGGGAAAGAGGTGTACCATCAGGAACATCACTAAACAGACCAGCCAGTCAAAAATAGATA | 1099 |
| Qy | 942 | tcacacgtaaaaaaatgcggggctctgagaagtccattactatcctctctactcctg | 1001 |
| Db | 1100 | TTCACCGTAAGAAATGACAGGTGTCAGAGAAACCGATTACAAATTCACTCAACCCCTG | 1159 |
| Qy | 1002 | aagcacctctgcgctgttaagtctattcttgagatattatgcataaaggaaagctcaagata | 1061 |
| Db | 1160 | AAGGCTGCTCAGCTGCATGTAAGATCAATTATGGAGATCATCGACAAGGAGCTCAGGATA | 1219 |
| Qy | 1062 | taaaattccacagaagagatccoccttgaagatttttagctcataatactttgttggacgtc | 1121 |
| Db | 1220 | CCAAGTCTCACTGAAGAAATCCCGCTTAAAAAATATTAGCACATAACAACATTTGTTGGCGCTC | 1279 |
| Qy | 1122 | tattggtaaaaggaagaatacttaaaaaattggacaagacacagacactaaataca | 1181 |
| Db | 1280 | TTATTGTTAAGAAGGAAGGAACCTTTAGAAANTTTGACGAAGATACAGATACAAAAATCA | 1339 |
| Qy | 1182 | cgatatctccattgcaggaattgacgtgtataatccagaacgcactattacagtttaag | 1241 |
| Db | 1340 | CAATATCTCCACTACAGGACTTGACACTGTACAATCCAGAAGCGACCATTTACAGTAAAG | 1399 |
| Qy | 1242 | gcaatgttgacatctgcacaagctggaggagagatcatgaagaaatacaggagctctt | 1301 |
| Db | 1400 | GCAGCATTTAGCCATGTGCAAGCGCGAAGAGAAATCATGAAAAAAATTTAGGGAGTCTCT | 1459 |
| Qy | 1302 | atgaaatgatattgctctatgaatctccaagcacatttaattccctggattaaatctga | 1361 |
| Db | 1460 | ATGAAATGATTTGCTGCTATGAATCTCCAGGCACATTTGATTCTCTGGTTTGAACCTGA | 1519 |
| Qy | 1362 | acgccttgggtctgttcccaccacttc-----agggatgccacctcccacctcagggc | 1415 |
| Db | 1520 | ATGCATTTGGGCCCTTTTCCCATCATCCTCATCAGGAATGCCACCACCTTCTGTTGGAGTTC | 1579 |
| Qy | 1416 | cccttcagcatgaactcctcctccaccgcagtttg----agcaatcagaacaggagactg | 1472 |
| Db | 1580 | CCTCACTACTCATCTACTTCTTATCCACCATTTTGGCGACGACGCCAGAGTTCAGAGACCG | 1639 |
| Qy | 1473 | tctacagtttaccagctctctatcagtcggtgcctcatctcgcaagcagggccagacaca | 1532 |
| Db | 1640 | TTCACTCTTCATCCAGCTTTTAGCTGTGGGAGCGATTTATTGGAAAGCAAGGACACACCA | 1699 |
| Qy | 1533 | tcaagcagctttctcgtcttctgtagcttcaattaaagtctccagcgcggaagcaccag | 1592 |
| Db | 1700 | TCAAAACAACCTATCACGCTTTTGCAGGAGCATCTATTAAAGTTGCACCTGCGGAGGACCTG | 1759 |
| Qy | 1593 | atgctaaagtggagtagtgatataactcgaacaccagaggtcagttcgaaggctcaggt | 1652 |
| Db | 1760 | ATGCCAACTAAGAATGGTGTATCATCTACTTGACCTCTCTGAAGCCCAATTTAAGGCTCAAG | 1819 |

| | | | |
|---|------|--|------|
| Qy | 1653 | gaagaatttatgaaaaaataaagaagaacacatttggtagtcctaagaagaagtgaaac | 1712 |
| Db | 1820 | GAAGGATCTATGTTGTAACCTTGAAGAAGAAAAAATCTTCGAGACCTTAAGAAGAAGTGAAC | 1879 |
| Qy | 1713 | ttagagctcatatcagagtgccatcccttggctgctggcagagttattggaaaagagagca | 1772 |
| Db | 1880 | TTGAGACCCCATTTAAAGTGCCTCATATGCTGCTGCAGCGGTTATTGTCGAAGAGGAA | 1939 |
| Qy | 1773 | aaacggtgaatgaacttcacgaatttgcagagtcagaagttgtgtccctcgtgaccaga | 1832 |
| Db | 1940 | AAACAGTAAATGAACCTTCAGAAATTAACAAGTCGAGAGGTTGTTGTCGCCGTGATCAA | 1999 |
| Qy | 1833 | cacctgatgagaatgacccaagtggtgtgcaaaaataactggctcactcttatgctgcagg | 1892 |
| Db | 2000 | CACCAGATGAAATGATCAAGTGGTCGTTAAGATAACGGGTCTACTTCTATGCAAGTCAGC | 2059 |
| Qy | 1893 | ttgccagagaataaattcaggaattctgactcaggttaa-----agcagcaccacaac | 1946 |
| Db | 2060 | TTGCACAAGGAAATTCAGAAATATTGGCTCAGGTAAAGAACAGACAGCACACACAGC | 2119 |
| Qy | 1947 | agaaggtctgcaagtggaacctcagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaag | 2005 |
| Db | 2120 | AGAAACAGTGCAAGTGGACAAACCCCAACCGAGAGAAAAATAAAGGTTCCAGGAAATGG | 2179 |
| Qy | 2006 | ccacacacagagcgagatgccaaacaaagacagagattgcttaaccacacagatggcgctg | 2065 |
| Db | 2180 | CCCTTCACAGACAGATGCCAGACCAACAGACACTTGAGATGAAGACAAATTTCCACTTC | 2239 |
| Qy | 2056 | accacctatccgaatcacatgcacagtttttacctagccagttgtttcttgaggaccag | 2125 |
| Db | 2240 | TTGTCCTCTGTTATGTGG-----GAAGAAATACCACTAGCGCATCCCCGTGTGAGGACCAG | 2295 |
| Qy | 2126 | gcaacttttgacctctgctctgtgagaatgtatatacttctctctctgaaatgtatga | 2185 |
| Db | 2296 | GCAACCATTAACAGT-----TCCTGAGATGATATTATTAGGCTTTCAAACTTTTGAA | 2350 |
| Qy | 2186 | caccagcgttttaaacaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaa | 2224 |
| Db | 2351 | ACCCAGCTTTAAAAATGCAAAAAAGGCAAAAAAGGAA | 2389 |
| RESULT 13 | | | |
| AF055923 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| African clawed frog. | | | |
| Xenopus laevis | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; | | | |
| Xenopodinae; Xenopus. | | | |
| 1 (bases 1 to 2161) | | | |
| Deshler,J.O., Hightett,M.I., Abramson,T. and Schnapp,B.J. | | | |
| A highly conserved RNA-binding protein for cytoplasmic mRNA | | | |
| localization in vertebrates | | | |
| Curr. Biol. 8 (9), 489-496 (1998) | | | |
| JOURNAL | | | |
| MEDLINE | | | |
| 98228351 | | | |
| 2 (bases 1 to 2161) | | | |
| Deshler,J.O., Hightett,M.I., Abramson,T. and Schnapp,B.J. | | | |
| AUTHORS | | | |
| Direct Submission | | | |
| TITLE | | | |
| Submitted (26-MAR-1998) Cell Biology, Harvard Medical School, 140 | | | |
| JOURNAL | | | |
| Longwood Ave., Boston, MA 02115, USA | | | |
| FEATURES | | | |
| source | | | |
| 1. .2161 | | | |
| /organism="Xenopus laevis" | | | |
| /db_xref="taxon:8355" | | | |
| 135..1919 | | | |
| /codon_start=1 | | | |
| /product="RNA binding protein Vera" | | | |
| /protein_id="AAC41285.1" | | | |
| /db_xref="GI:3136158" | | | |
| CDS | | | |


```
QY 2078 gaatcacatgcacagtttttaccttagccagttgtttctgaggagaccaggaactttttaa 2137
    |||
Db 2008 TGTGG---GAAGAAATACCACCTAGCGCATCCCTGTTTGTAGGACCGACGCAACCATTA 2063
    |||
QY 2138 ctctctctctgagaatgtatacttattctctctgtaaaatgtatgacacccacttita 2197
    |||
Db 2064 CTG-----TCTCTGGAATGATTATTTAGGCTTTCAAACCTTTTGAACCCAGCTTTAA 2118
    |||
QY 2198 aaacaaacaaacaaacaaacaa 2224
    |||
Db 2119 AAATGGAAAAAGGCAAAAAAGAAA 2145

RESULT 14
AF161270      2011 bp      mRNA          VRT      27-JUL-1999
LOCUS        Danio rerio Vg1 RNA binding protein mRNA, complete cds.
DEFINITION  AF161270
ACCESSION   AF161270.1  GI:5596631
KEYWORDS    zebrafish.
SOURCE      Danio rerio
ORGANISM    Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE   1 (bases 1 to 2011)
AUTHORS    Zhang,Q., Yaniv,K., Oberman,F., Wolke,U., Cit,A., Fromer,M.,
Taylor,W., Meyer,D., Standart,N., Raz,E. and Yisraeli,J.K.
TITLE      Vg1 RBP Intracellular distribution and evolutionarily conserved
expression suggest multiple roles during development
JOURNAL    Mech. Dev. (1999) In press
AUTHORS    Oberman,F., Fromer,M. and Yisraeli,J.K.
TITLE      Direct Submission
JOURNAL    Submitted (22-JUN-1999) Anatomy and Cell Biology, Hebrew University
Hadasah Medical School, POB 12272, Jerusalem 91120, Israel
LOCATION/Qualifiers
FEATURES             1..2011
source              /organism="Danio rerio"
                    /db_xref="taxon:7955"
                    /chromosome="linkage group 19"
                    /map="between Z3023 and Z15451"
                    1..1749
                    /note="Vg1 RBP"
                    /codon_start=1
                    /product="Vg1 RNA binding protein"
                    /protein_id="AAD45610.1"
                    /db_xref="GI:5596632"
                    /translation="MNKLYIGNVSEQASALDESIFEQWKIPFLVKSGYAFVDC
                    PDEKVMRAIDTLGKVLHGKLVLEHVSVPKRSCKLOIRNIPPHMOWEVLGDLA
                    QYGVESCEQVNTDETAVNVVRVYGAKDQAREMDKLNGLFMYNALKVSIPIETAA
                    ADAFAVGGRRGNFRPGPSLIGARKPLQSDVPDLKLVPTQFVGAILKEGATIR
                    NITQTHSKIDIHREKNAAGAEKPIVHSTPEGSSACRNIMEIQKEADIKTEEI
                    PLKILAHNNFVRLIGREKRLKIEQDTDKITISPLQDLTLNPERTIVKGTFLDA
                    CKAEEETMKVRESYENDVAAMHLSNIPGLNLNLPFGPAAGSGISPSVYSGPP
                    PGAQGYQSGAQMSESVHLFIPALAVGALIGKQGHILKLSRFAGASIKIADGI
                    DAKRMVISPPEAQKFAQGRIFGKLEENFPFKVEVKLEAHIKVPSFAAGRVIGK
                    GKTVELNLQNTSAEYVVPVDPQTFDENQVVVVKITGHFYASQLAQRKIQETISQVRQ
                    QQPRPSAAGPVPARRK"
BASE COUNT      581 a 494 c 501 g 435 t
ORIGIN
```

Query Match 20.2%; Score 845.8; DB 5; Length 2011;
Best Local Similarity 69.2%; Pred. No. 2.1e-160;
Matches 1220; Conservative 0; Mismatches 522; Indels 21; Gaps 4;

```
QY 251 atgaacaaactgtatatcggaacacctagcaggaacagcgccccctcggaacctagaaagt 310
    |||||
Db 1 ATGAATAAGCTATCATCGGAAATGTAGCGAGCGAGCGAGCGCTGGACTTGGAAAGT 60
    |||||
QY 311 atcttcaaggacgcaagaatcccgtgtcgggacccttccctggtgaagactgctacgcg 370
    |||||
```

Db 1135 GCGAGTGGAGGATTTCACCCCTCTGTGGTATCAGGACCACCGCCCTGGGGCTCAAGCTGGA 1194
 Qy 1436 cctaccocgcagtttgagcaatcagaacaggagagctgttcacatgatttccccagctcta 1495
 Db 1195 TATCAATCATTTGGGGCGCAGATGAGTTCGGAACCTGTGCACCTCTTCATTCTCTGCACTG 1254
 Qy 1496 tcaatcggtgccatcatcggaacagggccagcacatcacgaagctttctcgctttgct 1555
 Db 1255 GCAGTTGGTGTATTATCGGCAACGAGCGGCCAACACATTAAACAGCTCAGCGCTTCGCT 1314
 Qy 1556 ggagcttcaataagattgtctccagcggaagcaccagctgctaaagtggagatggtgatt 1615
 Db 1315 GGTCCGTCAATCAAAATGGCCCTGCTGTATGGAATTGATGCCAAGCAAGATGGTTATT 1374
 Qy 1616 atcactggaccacagaggtctcagttcaaggctcagggaagaatttatgaaaaattaaa 1675
 Db 1375 ATTTCTGGACCACAGAGGCACAGTTTAAAGGCTCAGGAGCTATCTTTGGAAAGTTAAAA 1434
 Qy 1676 gaagaaactttgttagtctaaagaagagtgaaacttgaagctctatcacagagtcca 1735
 Db 1435 GAGGAGAACCTTCTTCGGACCGAAAGAGAGGTTAAATTAGAGGCTCATATAAAAGTGCCA 1494
 Qy 1736 tccttctgctgagcagattattgaaaaggagcgaacggcggaatgaacttcagaat 1795
 Db 1495 TCATTCCGCTGCTGCAGAGTCATTGGTAAAGGAGGCAAAACGGTGAACGACTACAGAAC 1554
 Qy 1796 ttgcaagtcgagaagttgttgcctctgtaccagacacctgatgagaatgaccaaagt 1855
 Db 1555 TTGACCACTGGGAGGTGGTGTCTCCACGGGACCAGACACCTGTATGAATAATGACCAAGTA 1614
 Qy 1856 gtgtcaaaaactgtcacttctatgttgcaggttcccgagagaaaaattcagaaa 1915
 Db 1615 GTAGTAAATATCAGAGGCCACTTCTATGCAAGCCAGCTGGCCCGCAGAGAGATCCAGGAG 1674
 Qy 1916 attctgactcaggttaagcagcagcacaacagaagctctgcaaaagtggaccacctcag 1975
 Db 1675 ATATCTCTCAGGTGAGCGGCGACACAGCCCGCCCTCAGCCGCGAGTCCACCACTC 1734
 Qy 1976 tcaagcgggaagtaaggtctcag 1998
 Db 1735 GCCAGGAGAAATAGAGCCACTG 1757

RESULT 15
 AC023375/c
 LOCUS
 DEFINITION Homo sapiens chromosome 8 clone RP11-571M6 map 8, WORKING DRAFT
 SEQUENCE, 37 unordered pieces.
 AC023375
 AC023375.2 GI:7209933
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 158105)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Collangelo,M., Collins,S., Collumore,A., Cooke,P.,
 DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
 Fenestor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
 Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
 McEwan,P., McGurk,A., McKernan,C., McPheeters,R., Meldrum,J.,
 Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,

TITLE
JOURNAL

COMMENT

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
 Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,
 Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
 Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
 Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
 Zody,M.

Direct Submission

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 8, 2000 this sequence version replaced gi:6970513.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4227

Center clone name: 571_M_6

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 133946 bases at least Q40

Consensus quality: 143828 bases at least Q30

Consensus quality: 149412 bases at least Q20

Insert size: 174000; agarose-fp

Insert size: 154505; sum-of-contigs

Quality coverage: 3.2 in Q20 bases; agarose-fp

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 37 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1036: contig of 1036 bp in length

* 1037 1136: gap of 100 bp

* 1137 2213: contig of 1077 bp in length

* 2214 2313: gap of 100 bp

* 2314 3566: contig of 1253 bp in length

* 3567 3666: gap of 100 bp

* 3667 4804: contig of 1138 bp in length

* 4805 4904: gap of 100 bp

* 4905 6404: contig of 1500 bp in length

* 6405 6504: gap of 100 bp

* 6505 8366: contig of 1862 bp in length

* 8367 8466: gap of 100 bp

* 8467 10126: contig of 1660 bp in length

* 10127 10226: gap of 100 bp

* 10227 11460: contig of 1234 bp in length

* 11461 11560: gap of 100 bp

* 11561 13697: contig of 2137 bp in length

* 13698 13797: gap of 100 bp

* 13798 15334: contig of 1537 bp in length

* 15335 15434: gap of 100 bp

* 15435 16434: contig of 1000 bp in length

* 16435 16534: gap of 100 bp

* 16535 18900: contig of 2365 bp in length

* 18901 19000: gap of 100 bp

* 19001 20918: contig of 1918 bp in length

* 20919 21018: gap of 100 bp

* 21019 23110: contig of 2092 bp in length

* 23111 23210: gap of 100 bp

* 23211 25903: contig of 2693 bp in length

* 25904 26003: gap of 100 bp

```
* 26004 29042: contig of 3039 bp in length
* 29043 29142: gap of 100 bp
* 29143 32196: contig of 3054 bp in length
* 32197 32296: gap of 100 bp
* 32297 35269: contig of 2973 bp in length
* 35270 35369: gap of 100 bp
* 35370 39033: contig of 3662 bp in length
* 39032 39131: gap of 100 bp
* 39132 42030: contig of 2899 bp in length
* 42031 42130: gap of 100 bp
* 42131 46601: contig of 4471 bp in length
* 46602 46701: gap of 100 bp
* 46702 51387: contig of 4686 bp in length
* 51388 51487: gap of 100 bp
* 51488 55227: contig of 3740 bp in length
* 55228 55327: gap of 100 bp
* 55328 59146: contig of 3819 bp in length
* 59147 59246: gap of 100 bp
* 59247 62533: contig of 3293 bp in length
* 62540 62639: gap of 100 bp
* 62640 66443: contig of 3804 bp in length
* 66444 66543: gap of 100 bp
* 66544 72068: contig of 5525 bp in length
* 72069 72168: gap of 100 bp
* 72169 78427: contig of 6259 bp in length
* 78428 78527: gap of 100 bp
* 78528 83366: contig of 4839 bp in length
* 83367 83466: gap of 100 bp
* 83467 89726: contig of 6260 bp in length
* 89727 89826: gap of 100 bp
* 89827 97979: contig of 8153 bp in length
* 97980 98079: gap of 100 bp
* 98080 106168: contig of 8089 bp in length
* 106169 106268: gap of 100 bp
* 106269 114387: contig of 8119 bp in length
* 114388 114487: gap of 100 bp
* 114488 121719: contig of 7232 bp in length
* 121720 121819: gap of 100 bp
* 121820 130088: contig of 8269 bp in length
* 130089 130188: gap of 100 bp
* 130189 139207: contig of 9019 bp in length
* 139208 139307: gap of 100 bp
* 139308 158105: contig of 18798 bp in length.
FEATURES
  source
    1. 158105
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /map="8"
      /clone="RP11-571M6"
      /clone_lib="RPC1-11 Human Male BAC"
      /map="8"
    1. 1036
      /note="assembly_fragment"
      1137. 2213
      /note="assembly_fragment"
      2314. 3566
      /note="assembly_fragment"
      3667. 4804
      /note="assembly_fragment"
      4905. 6404
      /note="assembly_fragment"
      6505. 8366
      /note="assembly_fragment"
      8467. 10126
      /note="assembly_fragment"
      10227. 11460
      /note="assembly_fragment"
      11561. 13697
      /note="assembly_fragment"
      13798. 15334
      /note="assembly_fragment"
      15435. 16434
      /note="assembly_fragment"
```

```
misc_feature 16535. 18900
              /note="assembly_fragment"
misc_feature 19001. 20918
              /note="assembly_fragment"
misc_feature 21019. 23110
              /note="assembly_fragment"
misc_feature 23211. 25903
              /note="assembly_fragment"
misc_feature 26004. 29042
              /note="assembly_fragment"
misc_feature 29143. 32196
              /note="assembly_fragment"
misc_feature 32297. 35269
              /note="assembly_fragment"
misc_feature 35370. 39031
              /note="assembly_fragment"
misc_feature 39132. 42030
              /note="assembly_fragment"
misc_feature 42131. 46601
              /note="assembly_fragment"
misc_feature 46702. 51387
              /note="assembly_fragment"
misc_feature 51488. 55227
              /note="assembly_fragment"
misc_feature 55328. 59146
              /note="assembly_fragment"

Query Match 20.2%; Score 842.8; DB 2; Length 158105;
Best Local Similarity 88.8%; Pred. No. 7.5e-160;
Matches 847; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 1885 ttgccaggttccccagagaaaaattcaggaaattctgactcaggtaaaagcagcaccaca 1944
    ||| |
Db 9309 TTTCTAGGTGCTGCAAGAGTGACCACTCAGGAATTCGACTCAGGTAAACAGCACCACA 9250

Qy 1945 acagaagctctgaaagtgaccacctcagtcagacggaagtaaaagctcaggaaaa 2004
    ||| |
Db 9249 ACAGAAGGCTCTGCAAGAGTGACCACTCAGTCAAGACGGAAGTAAAGGCTCAGGAACA 9190

Qy 2005 gccaccacagagcagatgccaaacccaaagacagattgcttaaccaacagatggcgct 2064
    ||| |
Db 9189 GCCCACCACAGAGGAGATGCCAAACCAAGACAGATTGCTTAAACCAACAGATGGCGCT 9130

Qy 2065 gacccctatccagaatcacatgcacaagttttacctagccagttgttctgaggacca 2124
    ||| |
Db 9129 GACCCCTATCCAGATCACATGCACAAGTTTTCCTAGCCAGCTGTTTCTGAGGACCA 9070

Qy 2125 ggcaactttgaaactcctgtctctgtgagaa tgtatactttatgctctctgaaatgtatg 2184
    ||| |
Db 9069 GGCAACTTTTGAACCTCTGTCTGTGAGAAATGTATCTTTATGCTCTCTGAAATGTATG 9010

Qy 2185 acaccagctttaaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 2244
    ||| |
Db 9009 ACACCCAGCTTTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 8950

Qy 2245 gagaagagctctgcacttcccttctgtgtagctcacagataacagatatcttaattct 2304
    ||| |
Db 8949 GAGAAGAGCTCTGCACCTTCCCTTTGTTGTAGTCTCACAGTATAACAGATATCTAATTCT 8890

Qy 2305 tottaattcccccataatgccagaaaattggcttaagtatgctttcactaaattcatca 2364
    ||| |
Db 8889 TCTTAATATTCCTCCATAATATGCCAGAAATTTGGCTTAATGATGCTTCACTAAATTCATCA 8830

Qy 2365 aatagattgctcctaataatgccaaattgttaaaattggatcagaataattatcaacaggaactt 2424
    ||| |
Db 8829 AATAGATTGCTCCTAAATCCAAATTTGTTAAATTTGGATCAGATAATTAATTAACAGGAACCT 8770

Qy 2425 aaattgtaagccattagcatagaaaaactgttctcagtttttttttttttacttaacactaac 2484
    ||| |
Db 8769 AAATGTTAAGCCATTAGCATAGAAAAACCTGTTCTCAGTTTTATTTTACCTTACACTAAC 8710

Qy 2485 atgagtaacctaaaggaagtgctgaatggtgttgagggggtattaaacgttcattttta 2544
    ||| |
```

[illegible]

Search completed: January 25, 2002, 22:49:36
Job time: 28513 sec

| Result No. | Query | | | DB | ID | Description |
|------------|-------|-------|--------|----|-------------------|---------------------|
| | Score | Match | Length | | | |
| 1 | 4143 | 99.1 | 4159 | 4 | US-09-061-709-4 | Sequence 4, Appli |
| 2 | 804.4 | 19.2 | 2224 | 4 | US-09-261-855-1 | Sequence 1, Appli |
| 3 | 710.4 | 17.0 | 3412 | 4 | US-09-061-709-6 | Sequence 6, Appli |
| 4 | 695.6 | 16.6 | 3283 | 4 | US-09-061-709-8 | Sequence 8, Appli |
| 5 | 672.6 | 16.1 | 1946 | 4 | US-09-061-709-7 | Sequence 7, Appli |
| 6 | 580.6 | 13.9 | 1708 | 4 | US-09-061-709-5 | Sequence 5, Appli |
| 7 | 48.8 | 1.2 | 7218 | 1 | US-08-232-463-14 | Sequence 14, Appli |
| 8 | 44.8 | 1.1 | 518 | 1 | US-08-485-284A-2 | Sequence 2, Appli |
| 9 | 44.8 | 1.1 | 14796 | 4 | US-08-975-080-35 | Sequence 35, Appli |
| 10 | 44.8 | 1.1 | 14796 | 4 | US-09-630-706-10 | Sequence 10, Appli |
| 11 | 43.8 | 1.0 | 665 | 2 | US-08-883-795A-36 | Sequence 36, Appli |
| 12 | 41.8 | 1.0 | 7218 | 1 | US-08-232-463-14 | Sequence 14, Appli |
| 13 | 41.6 | 1.0 | 5852 | 1 | US-07-867-106-2 | Sequence 2, Appli |
| 14 | 41.2 | 1.0 | 5852 | 1 | US-07-867-106-2 | Sequence 2, Appli |
| 15 | 40.8 | 1.0 | 742 | 1 | US-07-847-010-12 | Sequence 12, Appli |
| 16 | 40.8 | 1.0 | 1562 | 1 | US-08-330-108-1 | Sequence 1, Appli |
| 17 | 40.8 | 1.0 | 1562 | 5 | PCR-US92-10087-1 | Sequence 1, Appli |
| 18 | 40.6 | 1.0 | 4159 | 4 | US-09-061-709-4 | Sequence 4, Appli |
| 19 | 40.4 | 1.0 | 6243 | 2 | US-09-056-075-1 | Sequence 1, Appli |
| 20 | 40.2 | 1.0 | 1758 | 3 | US-08-836-567-3 | Sequence 3, Appli |
| 21 | 39.8 | 1.0 | 2726 | 1 | US-08-281-714-1 | Sequence 1, Appli |
| 22 | 39.8 | 1.0 | 37950 | 4 | US-09-338-907-183 | Sequence 183, Appli |
| 23 | 39.6 | 0.9 | 4140 | 3 | US-08-894-731-2 | Sequence 2, Appli |
| 24 | 38.8 | 0.9 | 4235 | 1 | US-08-021-601-3 | Sequence 3, Appli |
| 25 | 38.8 | 0.9 | 4235 | 1 | US-08-082-849B-3 | Sequence 3, Appli |
| 26 | 38.8 | 0.9 | 4235 | 5 | PCR-US94-01624-3 | Sequence 3, Appli |
| 27 | 38.8 | 0.9 | 51952 | 3 | US-08-947-823-1 | Sequence 1, Appli |

QY 301 cctagaagaatcttcaagagacccaagatcccgggtgtcgggaccttctcgtggaagac 360
|||||
Db 301 cctagaagaatcttcaagagacccaagatcccgggtgtcgggaccttctcgtggaagac 360
|||||
QY 361 tggctacgcgttcgtggaactgcgcgacagagactgggacctcaagggccaatcgagcgct 420
|||||
Db 361 tggctacgcgttcgtggaactgcgcgacagagactgggacctcaagggccaatcgagcgct 420
|||||
QY 421 ttcaggtaaaatagaactgcacgggaaaccca tagaagtttgagcactcggctcccaaaaag 480
|||||
Db 421 ttcaggtaaaatagaactgcacgggaaaccca tagaagtttgagcactcggctcccaaaaag 480
|||||
QY 481 gcaaaagattcggaaacttcagatacgaaata tccgcgtcatttacagtggaagtgct 540
|||||
Db 481 gcaaaagattcggaaacttcagatacgaaata tccgcgtcatttacagtggaagtgct 540
|||||
QY 541 ggtatgttactagtcagtatgagtggtggagagctgtggaagctgtggaagtgaaactgactc 600
|||||
Db 541 ggtatgttactagtcagtatgagtggtggagagctgtggaagctgtggaagtgaaactgactc 600
|||||
QY 601 ggaactgcagttgtaaatgaacctattccagtaaggagccaagctagacaagcactaga 660
|||||
Db 601 ggaactgcagttgtaaatgaacctattccagtaaggagccaagctagacaagcactaga 660
|||||
QY 661 caaactgaatggatttcagttagagaatttcaaccttgaaagtgcctatatccctgatga 720
|||||
Db 661 caaactgaatggatttcagttagagaatttcaaccttgaaagtgcctatatccctgatga 720
|||||
QY 721 aatggccgcacgaacaaacccctgcagcagcccgagctcgcggggcttgggcagag 780
|||||
Db 721 aatggccgcacgaacaaacccctgcagcagcccgagctcgcggggcttgggcagag 780
|||||
QY 781 gggctcctcaaggcaggggtctccaggatccgtatccaagcagaacccaatgtgatttgc 840
|||||
Db 781 gggctcctcaaggcaggggtctccaggatccgtatccaagcagaacccaatgtgatttgc 840
|||||
QY 841 tctgcgcctgctgggttcccaacccaattgttgagccaatcattaggaaaagagtgccac 900
|||||
Db 841 tctgcgcctgctgggttcccaacccaattgttgagccaatcattaggaaaagagtgccac 900
|||||
QY 901 catcggaaacatcaccaaacagaccagctcaaaatcgatgtccacgtaaaagaaatgc 960
|||||
Db 901 catcggaaacatcaccaaacagaccagctcaaaatcgatgtccacgtaaaagaaatgc 960
|||||
QY 961 gggggctgtcgaagaatcgattactatcctctactcctcctggaaggcacctctcggcttg 1020
|||||
Db 961 gggggctgtcgaagaatcgattactatcctctactcctcctggaaggcacctctcggcttg 1020
|||||
QY 1021 taagtctattctggagattatgcataaggaagctcaagata taaaattcacagaagagat 1080
|||||
Db 1021 taagtctattctggagattatgcataaggaagctcaagata taaaattcacagaagagat 1080
|||||
QY 1081 ccccttgaagattttagctcataataacttgttgagctcttatttggtaaaagaaag 1140
|||||
Db 1081 ccccttgaagattttagctcataataacttgttgagctcttatttggtaaaagaaag 1140
|||||
QY 1141 aaatcttaaaaaaatttgacaagacacagacactaaaatacagatattcctcattgcaga 1200
|||||
Db 1141 aaatcttaaaaaaatttgacaagacacagacactaaaatacagatattcctcattgcaga 1200
|||||
QY 1201 attgagcgtgtataaaccgaagacgactattacagttaaaggcaatgtgtgagacatgtgc 1260
|||||
Db 1201 attgagcgtgtataaaccgaagacgactattacagttaaaggcaatgtgtgagacatgtgc 1260
|||||
QY 1261 caaagctgagaggagatcatgaagaaaatcaggggactcttagaaaatgatatgtcttc 1320
|||||
Db 1261 caaagctgagaggagatcatgaagaaaatcaggggactcttagaaaatgatatgtcttc 1320
|||||
QY 1321 tatgaatcttcaagcactttaaattcctggattaaatctgaacgccttgggtgttccc 1380
|||||
Db 1321 tatgaatcttcaagcactttaaattcctggattaaatctgaacgccttgggtgttccc 1380
|||||
QY 1381 acccaacttcagggtatgccacctcccacctcaggggcccccttcagccactgactcctccta 1440
|||||

Db 1381 acccaacttcagggtatgccacctcccacctcaggggcccccttcagccactgactcctccta 1440
|||||
QY 1441 ccgcagtttgagaatacagaacgagagactgttcacagttttatccacagctctatcagt 1500
|||||
Db 1441 ccgcagtttgagaatacagaacgagagactgttcacagttttatccacagctctatcagt 1500
|||||
QY 1501 cgtgcacatcatcggcaagcagggccagcagcatcaagcagctttctcgttttctgaggc 1560
|||||
Db 1501 cgtgcacatcatcggcaagcagggccagcagcatcaagcagctttctcgttttctgaggc 1560
|||||
QY 1561 ttaaatgaagattgcttccagcgggaagcaccagatgctaaagtggagatggtgattatcac 1620
|||||
Db 1561 ttaaatgaagattgcttccagcgggaagcaccagatgctaaagtggagatggtgattatcac 1620
|||||
QY 1621 tggaccacagagctcagttcaagctcaggggaagaatttatggaaaaataaagaaga 1680
|||||
Db 1621 tggaccacagagctcagttcaagctcaggggaagaatttatggaaaaataaagaaga 1680
|||||
QY 1681 aaacttggtagtctctaaagaagagtgaaacttgaagctcatcatcagagtgccactctt 1740
|||||
Db 1681 aaacttggtagtctctaaagaagagtgaaacttgaagctcatcatcagagtgccactctt 1740
|||||
QY 1741 tggctgtgcagagtttatggaaaaagggcaaaaacggtgaatgaacttcagaaatttgc 1800
|||||
Db 1741 tggctgtgcagagtttatggaaaaagggcaaaaacggtgaatgaacttcagaaatttgc 1800
|||||
QY 1801 aagtcgagaagttgtgtccctcgtgaccagacacctgtagagaatgaccaagtgggtgt 1860
|||||
Db 1801 aagtcgagaagttgtgtccctcgtgaccagacacctgtagagaatgaccaagtgggtgt 1860
|||||
QY 1861 caaaaataacttggtcaacttctatgcttgcagggttgcagagaaaaattcagagaaattct 1920
|||||
Db 1861 caaaaataacttggtcaacttctatgcttgcagggttgcagagaaaaattcagagaaattct 1920
|||||
QY 1921 gactcgggtaaagcagcaccacaacagaaaggtctgtcaaaagtggaccacctcagtcag 1980
|||||
Db 1921 gactcgggtaaagcagcaccacaacagaaaggtctgtcaaaagtggaccacctcagtcag 1980
|||||
QY 1981 acggaagtaaaagctcaggaacagcccaccacagagcgagatgccaacacaagacaga 2040
|||||
Db 1981 acggaagtaaaagctcaggaacagcccaccacagagcgagatgccaacacaagacaga 2040
|||||
QY 2041 ttgcttaacacaagatgggcgtgacccctcatccagaaatccacatgcacaagtctttac 2100
|||||
Db 2041 ttgcttaacacaagatgggcgtgacccctcatccagaaatccacatgcacaagtctttac 2100
|||||
QY 2101 ctgaccagttgttctgagaccaggaactttgaaactcctgtcctgtgagaagtgtat 2160
|||||
Db 2101 ctgaccagttgttctgagaccaggaactttgaaactcctgtcctgtgagaagtgtat 2160
|||||
QY 2161 actttatgctctgaaatgtagacccagcttttaaaacaacaaacaaacaaacaa 2220
|||||
Db 2161 actttatgctctgaaatgtagacccagcttttaaaacaacaaacaaacaaacaa 2220
|||||
QY 2221 aaaaaggtggggaggggaaagagaagagctgtgcacttcccttgtttagtctc 2280
|||||
Db 2221 aaaaaggtggggaggggaaagagaagagctgtgcacttcccttgtttagtctc 2280
|||||
QY 2281 cagtaataacagatatcttaattcttctaattccccataatgccagaaattggctta 2340
|||||
Db 2281 cagtaataacagatatcttaattcttctaattccccataatgccagaaattggctta 2340
|||||
QY 2341 atgatgtcttcaactaaattcatcaaatagattgtccttaaatccaattgttaaaattgga 2400
|||||
Db 2341 atgatgtcttcaactaaattcatcaaatagattgtccttaaatccaattgttaaaattgga 2400
|||||
QY 2401 tcagaataattatcacaggaacttaaatgtttaagccattagcatagaaaaactgtctc 2460
|||||
Db 2401 tcagaataattatcacaggaacttaaatgtttaagccattagcatagaaaaactgtctc 2460
|||||
QY 2461 gttttatttttaactaacataacatgataacctaagggaagtgctgtaagtgtgttgc 2520
|||||

Db 2461 gttttatttttaacctaaactaacatgaglaacctaaagggaagtgtgaatggtgtgac 2520
QY 2521 aggggtattaaacgtgcatttttaactcaactacctcaaggtattcagtaatacaatgaaaa 2580
Db 2521 aggggtattaaacgtgcatttttaactcaactacctcaaggtattcagtaatacaatgaaaa 2580
QY 2581 gcaaaatgttccctttttttttgaaaaattttatatactttataatgatagaagtcaccaacg 2640
Db 2581 gcaaaatgttccctttttttttgaaaaattttatatactttataatgatagaagtcaccaacg 2640
QY 2641 ttttttaaaaaataaatttaaaatttaaacagcaaatcaacagctaacagggcaaatgaatttt 2700
Db 2641 ttttttaaaaaataaatttaaaatttaaacagcaaatcaacagctaacagggcaaatgaatttt 2700
QY 2701 tacttctggctggtgacagtaaaagctggaaaatttaatttcagggttttttgaggttttg 2760
Db 2701 tacttctggctggtgacagtaaaagctggaaaatttaatttcagggttttttgaggttttg 2760
QY 2761 acacagttattagtttaaatcaaatgttcaaaatgaatcaggagcagtgcttagtctggaga 2820
Db 2761 acacagttattagtttaaatcaaatgttcaaaatgaatcaggagcagtgcttagtctggaga 2820
QY 2821 gcagcactaccatttattcttcattttatagttggaaaagtttttgacggtactaaca 2880
Db 2821 gcagcactaccatttattcttcattttatagttggaaaagtttttgacggtactaaca 2880
QY 2881 tgggtgcgaggagattttggaacgctggttttaaatgcttcagagacttcagtttttt 2940
Db 2881 tgggtgcgaggagattttggaacgctggttttaaatgcttcagagacttcagtttttt 2940
QY 2941 gtttagctacatgattggaatgcataataaattgctttgtcttctgactactcaatacctaa 3000
Db 2941 gtttagctacatgattggaatgcataataaattgctttgtcttctgactactcaatacctaa 3000
QY 3001 aqaaagtgcacagtgaaagatgaagacttcaactgactggcaaaaagcaagcttta 3060
Db 3001 aqaaagtgcacagtgaaagatgaagacttcaactgactggcaaaaagcaagcttta 3060
QY 3061 gctgtcttatagtagtcttagttgcccactacactcagacccaatggacagctacaga 3120
Db 3061 gctgtcttatagtagtcttagttgcccactacactcagacccaatggacagctacaga 3120
QY 3121 tgggtgacagtggtttaaacgcacaaagagctacatttccatggggccagcactgtcat 3180
Db 3121 tgggtgacagtggtttaaacgcacaaagagctacatttccatggggccagcactgtcat 3180
QY 3181 gagcctcaactaagctattttgagatttttaagcactgataaattaaaaaataaaaaa 3240
Db 3181 gagcctcaactaagctattttgagatttttaagcactgataaattaaaaaataaaaaa 3240
QY 3241 aaattagactccaccttaagttagtaagataacaggaatttctgtatactgtgcaatcag 3300
Db 3241 aaattagactccaccttaagttagtaagataacaggaatttctgtatactgtgcaatcag 3300
QY 3301 tttctttgaaaaaagtcacaaagatagagaatcacagaaaagtttngggatataatttg 3360
Db 3301 tttctttgaaaaaagtcacaaagatagagaatcacagaaaagtttngggatataatttg 3360
QY 3361 aatgactgtgaaaacatatgacctttgataacgaaactaatttgcactccttgacagca 3420
Db 3361 aatgactgtgaaaacatatgacctttgataacgaaactaatttgcactccttgacagca 3420
QY 3421 aagccagtagctacaattgtttgggtgtgggtgtgggtgtcccaagccagcgtgctctctg 3480
Db 3421 aagccagtagctacaattgtttgggtgtgggtgtgggtgtcccaagccagcgtgctctctg 3480
QY 3481 aattgattttttgagtttttgnntnaagatgatcacagncatgtttacactgattctnaa 3540
Db 3481 aattgattttttgagtttttgnntnaagatgatcacagncatgtttacactgattctnaa 3540
QY 3541 ggacatatntataacccttttaaaaaaaatccccctgcctcatcttatttccgagatga 3600
Db 3541 ggacatatntataacccttttaaaaaaaatccccctgcctcatcttatttccgagatga 3600

QY 3601 atttcgatcacagactagatgttcttctgaagatcaattagacatttntgaaaaatgatttaa 3660
Db 3601 atttcgatcacagactagatgttcttctgaagatcaattagacatttntgaaaaatgatttaa 3660
QY 3661 agtgtttctcttaattgtctctgaaaaaagtttctttttagtattttaaccacaaaagt 3720
Db 3661 agtgtttctcttaattgtctctgaaaaaagtttctttttagtattttaaccacaaaagt 3720
QY 3721 ccccttttgcactggtttctctcctagcattcatgatttttttttcacacaaatgaattaaa 3780
Db 3721 ccccttttgcactggtttctctcctagcattcatgatttttttttcacacaaatgaattaaa 3780
QY 3781 attgctaaaaatcatgactggtcttctggttgcatttcaggttaagatgtgtttlaaaggcca 3840
Db 3781 attgctaaaaatcatgactggtcttctggttgcatttcaggttaagatgtgtttlaaaggcca 3840
QY 3841 gagctttctcagttattgattttttttccccaataattgattttttaaaaaatatacacat 3900
Db 3841 gagctttctcagttattgattttttttccccaataattgattttttaaaaaatatacacat 3900
QY 3901 aggagctgatttaaaacacctgctggttttaaatctctcanatttcaactctagccttta 3960
Db 3901 aggagctgatttaaaacacctgctggttttaaatctctcanatttcaactctagccttta 3960
QY 3961 gtaggcnaatcanaatttacttttacttaagcatttgaatttggagtatctcgttacta 4020
Db 3961 gtaggcnaatcanaatttacttttacttaagcatttgaatttggagtatctcgttacta 4020
QY 4021 gtaagaataatcnaatgagtttcttactcnaaanaatgggtcattctcctcatgn 4080
Db 4021 gtaagaataatcnaatgagtttcttactcnaaanaatgggtcattctcctcatgn 4080
QY 4081 ataagttnccccaatgcagcttcattttccaganaaccttgacgcaggataaattttttc 4140
Db 4081 ataagttnccccaatgcagcttcattttccaganaaccttgacgcaggataaattttttc 4140
QY 4141 atcatttaggtccccaaaa 4159
Db 4141 atcatttaggtccccaaaa 4159

RESULT 2

US-09-261-855-1
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; FILE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/261,855A
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2224
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-261-855-1

Query Match 19.2%; Score 804.4; DB 4; Length 2224;
Best Local Similarity 67.0%; Pred. No. 6.9e-199;
Matches 1141; Conservative 0; Mismatches 561; Indels 0; Gaps 0;

QY 247 cacaatgaacacactgtatatcgaaacctcagcgagaaacgcgccttcggacactaga 306
Db 127 caccatgaacacagctttacatcggaacctcaacgagagtgctgacccccgcgacactgga 186
QY 307 aagatattctcaaggacgccaagaatcccggtgtcgagacccttctcgttggaagactggcta 366
Db 187 gaaagtattcgcggagcacaagatctcctacagcgccagttcttctggtcaaatccggcta 246

QY 367 cgcattcgtagctgcccggacagagagctgggcccctcaaggccatcagagcgcttttcagg 426
Db 247 cgccttcgtgattgcccgcagcagcactggcgatgaaggccatcgaactttctcggg 306
QY 427 taaatagaactgcacgggaaaccatagaaagtggagcactcgtcccaaaaggcaag 486
Db 307 gaaagtgaactgcaggaagaaacgtctagagattgaaacactcagtcaccaaaacaaag 366
QY 487 gattcgaaacttcagatagaaatatcccgcctcatttacagtgggaggtgctgggatag 546
Db 367 gattcgaaatacagatccgcaatttccacctcagctccgatgggaagtgcctagatag 426
QY 547 ttactagtcagatggagtgtggagagctgtgagcaagtgaacactgactcggaaac 606
Db 427 cctgctggctcagtcacgtggaagactgtgagcaagtgaacactgaaagtggagac 486
QY 607 tgcagttgaaattgaacctattccagtaaggacacagctagacacactagacaact 666
Db 487 agcgttggaactccaccttactctaacccgggagcagaccaggcaagctatcatgaagct 546
QY 667 gaattgatttcagtagaataattcaacttgaaagttagcctatatccctgatgaatggc 726
Db 547 aatggccatcaactggagaacatgcccgaaggtctctacatacctgatgagcagat 606
QY 727 cgcaccagaaaccccttcagcagaccccagagctgcgcgggggttgggagaggggctc 786
Db 607 aacacaaaggtcctgagaatggcgctgtggaggtttgggtctcggggccagccccgca 666
QY 787 ctcaaggcaggggtctccagatccgtatccaagcagaaacacatgtattgtcctcgg 846
Db 667 aggtctcccgtggcagcaggggtccagccaagcagcagccagtgacatccctctccg 726
QY 847 cctgctggttccaccacaaattgttgagccatcataggaagaaagggtgcacaccatcg 906
Db 727 gctcctggtgctacgcagtatgtaggcgtatcatcattggcaaggaggtgccaccatcg 786
QY 907 gaacatcaccaaaacagaccacagctaaatcgatgtccaccgttaaagaaatggggggc 966
Db 787 aaacatcacaaacagacgcagctccaaatagacgtgcataaggaaaggagatggcggc 846
QY 967 tgcagaaagtcgattactatcctctactcctcctgaaggccactctgcgcttgtagtc 1026
Db 847 tgcggaagggccatcagcgtgcatcaacccctgaaggctgctccctccgcgcgaagat 906
QY 1027 tattctggagattatgcataaggaagctcaagatatataatcacagaagagatccctt 1086
Db 907 gatctggagattatgcataaggggcaaggacacacaaacggcagatgaagtccctt 966
QY 1087 gaagattttagctcaataaactttgttgacgtctctatttggttaaagaaaggaaatct 1146
Db 967 gaagatcctggctcaatacaactctcgtcggcgactcattggcaaggagccggaacct 1026
QY 1147 taaaaaattgagcaagacacacacataaaatcacgatatctcattcgaggaattgac 1206
Db 1027 gaagaaagtgagcagacacagacagagatcaacatctcctcgtccagacacctac 1086
QY 1207 gctgataatccgaaacgcactattacagttaaaggcaatgttgagacatgtgccaagc 1266
Db 1087 gctataacccctgagaggaccatcactgtgaaggcgccatcagaaactgtgcagggc 1146
QY 1267 tgaggaagagatcatgaagaaatcaggggaggttattgaaaaatgatattgtctctatgaa 1326
Db 1147 cgagcagagatcatgaagaaagtctgagaggtcttaagaaagacgtggcgccatgag 1206
QY 1327 tctcaagcagcatttaattccttgattaaatcagacgccttggtctgtgtccaccac 1386
Db 1207 ctgtcagtcaccactcctcctgggttaacctggctcgtgtgaggtctctccagcttc 1266
QY 1387 ttcaggatgccacctccaccctcagggcccccttcagccatgactcctccctaccgcga 1446
Db 1267 atccagcgctgctccctcctccagcagtgctactgggggtgctccctctatagctcctt 1326

QY 1447 gttgagcaatcagaacacgagagactgttcatcagtttatccaccagctctatcagtgctg 1506
Db 1327 catgagctccggagcagcagagatggtacaagtgttcatcccccagcgtgtggggcg 1386
QY 1507 catcatcggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1566
Db 1387 catcatggcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1446
QY 1567 taagattgctccagcgggaagcaccagatgctaaagtggaggtggtgattatcactggacc 1626
Db 1447 caagattgctccaccagaaacacactgactccaagttcgaagtgtcgtcatcactggacc 1506
QY 1627 accagagctcagttcaagcgtcagggagaatatttgaaaaattaaagaagaaaaactt 1686
Db 1507 cccagagctcagttcaagcgtcagggagaatatttgcaaaactaaagaagaaattt 1566
QY 1687 tgttagtccctaaagcagaggtgaaactgaagctcatatcagagtgccatcctttgctg 1746
Db 1567 ctttgggtcccaagcaggaagttaaagctagagaccacatacgggttccgggttcagcagc 1626
QY 1747 tggcagagttattgaaaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1806
Db 1627 cggcgcgtcatcgcaaaaggcgcaaaacggtgaatgagctgcagaacttgactgcagc 1686
QY 1807 agaagtgtgtcctcgtgacacagacacctgtgagaatgaccaaagtgtgtgtcaaaat 1866
Db 1687 tgaggtgtgtgagcgaagacacagacccggatgagaagcacaagctattgttaagat 1746
QY 1867 aactggtcacttctatctgtcagaggtgcccagagaaaaattcagaaaattcagactca 1926
Db 1747 catgggacattctcgtccagcagatgggtccagcgaagatcccgagacatccctggctca 1806
QY 1927 ggtaaagcagcaccacaacacag 1948
Db 1807 agttaagcaacagcaccagaag 1828

RESULT 3
US-09-061-709-6
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; FILE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-6

Query Match 17.0%; Score 710.4; DB 4; Length 3412;
Best Local Similarity 64.5%; Pred. No. 2.1e-174;
Matches 1155; Conservative 0; Mismatches 556; Indels 81; Gaps 3;
QY 251 atgaacaactgtatatcggaacactcagcagagacgcgcccccctcgagacctagaagt 310
Db 73 atgaacaagctttacatcgggaacactgagcccgccgtcaccgcgcgacacctccgcag 132
QY 311 attctcaaggacgccaagatcccgggtgtcgggacctctcgtgtgaaagactggtctacgcg 370

Db 133 ctctttgggacaggaagctgccctggcggaagagtgctctgctgaagtccggctaagcc 192
QY 371 ttctggactgccgagagagctggccctcaaggccatcgagcgctttcaggtaaa 430
Db 193 ttctggactaccgacaggaactggccatccggccatcgagacctctcgggtaaa 252
QY 431 atagaactgcacgggaacccataagaattgagcactcgttcccaaaaagcaaggatt 490
Db 253 tgggaattgcatgggaatactcgtgaagtgtattactcagctctcaaaagctaaggagc 312
QY 491 cggaaactcagatcagaaatccgctcatttaccagtggagtgctgctgatatgta 550
Db 313 agaaaaattcagatcgaataatccctcctccctcgcagtggaggtgttgatgact 372
QY 551 ctagtccagtatggagtgtgagagctgtgagcaagtgaacactgactcggaaactgca 610
Db 373 ttggctcaatatggacagtggagaatgtggaacaagtcaacacagacagaaacccgc 432
QY 611 gtgttaaatgtaacctattccctgaaggaaccaagttagaagcactagacaaactgaat 670
Db 433 gtgtcaactgcacatatgcaacaagagaagaacaaatagccatggagaagctaaagc 492
QY 671 ggatttcagttagaaatttcaaccttgaagtgcgcgggggttcccgagagggctcctca 730
Db 493 gggcatcagttgagaactactccttcaagatttctcatcctccggatgaagagtgaagc 552
QY 731 cagcaaaaccccttgcagcagcccgaggtgcgcgggggttggcgagaggggtcctca 790
Db 553 tcccttgcctccctcagcgagcccgagcgtggggaccacttcccgggagcaagggccac 612
QY 791 agcgaggggttcccgaggtccgtatccaaagcagaaccatgtgatttgcctctgcgcctg 850
Db 613 g-----cccttggggcacttctcagggcagacagattgatttcccgctgcggatc 663
QY 851 cftggtcccaaccatattgttgagccatcatagaaagaagggtgcaccaattcggaaac 910
Db 664 cftgtcccaaccatattgttgggtgccatcatcggaaagggtgttgcacataaagaac 723
QY 911 atcaccaaacagaccagctctaaatcatgctccaccgttaaagaataatgcggggctgt 970
Db 724 atcataagcagaccagctcccggttagatatccatagaaaaagagaactcttggagctga 783
QY 971 gagaagtcgattactactctactctgaagcactctgcggcttgaagtctatt 1030
Db 784 gagaagctgtcaacctcatccaccocagagggactctgaagcatgcgcgatatt 843
QY 1031 cftggagattatcataaaggaactcaagataataaattcaagaagagatcccttgaag 1090
Db 844 cftgaaatcatgcagaaagagcgagatgagaccaaactagccgaagagattcctctgaa 903
QY 1091 attttagctcataataactttgttgacgtcttattggttaaagaaggaataatcttaa 1150
Db 904 atcttggcacacaatggcttgggttgaagactgattggaaaaagaaggcagaaatttgaag 963
QY 1151 aaaaattgagcaagacagacactaaaatcacgatatctccatttcaggaaattgacgtg 1210
Db 964 aaaaattgaactgaacaggggaccacaagataacaatctcatcttgcaggatttgacata 1023
QY 1211 tataatccagaacgcactattacagttaaaggcaatgttgagacatgtgccaaactgaag 1270
Db 1024 tacaacccgggaagaaccatcactgtgaaggggcagcagttgagggcctgtgccagtgctgag 1083
QY 1271 gaggagatcatgaagaaatccggagtgcttatagaataatgatatgtcttctatgaatt 1330
Db 1084 atagagattatgaagaagctgcgtgagggccttggaaatgatatgtctggtgttaaccaa 1143
QY 1331 caagcacatttaattccttgattaaatctgaacgccttgggtctgttcccaaccacttca 1390
Db 1144 caagccaattctgatacccgaggttgaacctcagcgcacttggcatcttttcaacaggaactg 1203
QY 1391 gggatgccacctcccaacctcagggccctcttcagcagatgaactcct----- 1435
Db 1204 tccgtgctatctcccaacagaggcccccggagctccccccgctgccccctaccacccc 1263

QY 1436 -----ccctaaccgcagtttgagc----- 1454
Db 1264 tttaactaaccactcoggaatacttctcagcctgtaccoccatcaccagtttggcggcttc 1323
QY 1455 -----aatcagaaacggagactgttcatcagtttcccaagcttatca 1498
Db 1324 cgcgatcatcactcttaccagacgagagattgtgaactcttctatcccaaccaggtc 1383
QY 1499 gtcggtgccatcatcgccaagcagggccacacatcaagcagcttctcgtcttgcctgga 1558
Db 1384 gtgggcgcctatcatcgggaagggggcacacatcaaacagctgcgcgagattcgccgga 1443
QY 1559 gtttcaattaaagtctccagcggggaagcaccagatgctaaaagtggagtggtgattatc 1618
Db 1444 gctctatcaagatgcccctcggaagggccacagcgtcagcgaaaggtatggtcatcctc 1503
QY 1619 actgaaccacccagaggtcagttcaaggctcagggggaagaatttatggaaaaataaagaa 1678
Db 1504 accgggcccacggaaagccagttcaaggccagggcgagatcttcttgggaaactgaaagag 1563
QY 1679 gaaaactttgttagtcttaaaagaaggtgaaacttgaagctcatatcagagtgcctatcc 1738
Db 1564 gaaaactctttaaaccacaaagaagtgaaagtggagctggaagcgcatatccagagtcctct 1623
QY 1739 ttgtgctggcagagttatttgaaaaaggaggcaaaacggtgaaatcgaacttcagaatttg 1798
Db 1624 tccacagctggcgggtgattggcaaaagtggaagcgcgtgaacgaactgcagaactta 1683
QY 1799 tcaagtgcagaagttgttgcctcctgcacagacacacctgatgagaatgaccaagtggtt 1858
Db 1684 accagtcgaagaatcatcgtcctcgtgaccaaaccagatgaaaatgaggaaagtgc 1743
QY 1859 gtcaaaataactggcacttctatcgttccaggttgcacagagaaaaattcaggaaatt 1918
Db 1744 gtcaaatattcggcacttcttctgtagcactgcacagcgcaagatcaggggaaatt 1803
QY 1919 ctgactcagttgaagcagcaccacaaacaggaaggtctctgcaaaagtggaccac 1970
Db 1804 gtacaacaggtgaagcagcagaggaagaataaccttcaggggagtcgcctcac 1855

RESULT 4
US-09-061-709-8
; Sequence 8, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-8

Query Match 16.6%; Score 695.6; DB 4; Length 3283;
Best Local Similarity 64.8%; Pred. No. 1.4e-170;
Matches 1115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;
QY 251 atgaacaaactgtatatcggaacctcagcgagaagcggccccctcggacctagaagt 310

```
Db 73 atgaacagctttacatcggaactgagccccccgctccacgcgcagacctccgcag 132
Qy 311 atctcaaggacgcaagatccccggttcggaaccccttctggtgaagactggctacgcg 370
Db 133 ctctttgggacaggaagctgccccctgcggaagctctgctggaagtcggtacgc 192
Qy 371 ttcgtgactgcccggacgagagctgggcccctcaagccatcgagggcgttctcaggtaaa 430
Db 193 ttcgtgactaccccgcacgaactggccatccgcgcacatcgagacctctcggtgtaa 252
Qy 431 atagaactgcacgggaacccatagaagttgagcactcgcgtcccaaaaagcaagatt 490
Db 253 gtgaattgcattgggaataatcgaagttgattcactcagctctcaaaaagctaaagagc 312
Qy 491 cggaaactcagatcacgaataatccccctcatttaccagtgggaggtgctgtagttaa 550
Db 313 aggaataatcagattcgaataatccccctcactcactgcagtgagggtgttgatggactt 372
Qy 551 ctagtccagtatgagtggtgagagctgtgagcaagtgaacactgactcggaaactgca 610
Db 373 ttggtcaaatgggacagtggaatgtggaacaagtcaacacagacacagaacccgcc 432
Qy 611 gttgtaaatcaactattccagtaagacacaagctagacaagcactagacaaactgaat 670
Db 433 gttgtcaagctcacatatgcacaaggaagaagcaaaaatagccaagggaagctaaagc 492
Qy 671 ggatttcagttagagaatttcaccttgaaagtagccctatatccctgatgaaatggcgcc 730
Db 493 gggcatcagtttgagaactactctcaagatttccctacatcccgatgaagagtgagc 552
Qy 731 cagcaaaaaccccttgacgagcccgagggctgcgggggttggcgagagggggtctctca 790
Db 553 tccccctgcgccccctcagcagcgcagcgtgggacacactcttcccgggagcaaggc--- 609
Qy 791 aggcaggggtctccagatcogtatccaaagcagaaacatgtgattgctctgcgctg 850
Db 610 -----cagccccctggggcaacttccagccagacagattgatttcccgcgtcgagtc 663
Qy 851 ctggttcccccccaatttggtagccatcataggaagaaagaggtgcacacattcggaaac 910
Db 664 ctggtcccccccaagtttggtagccatcatcggaagagggctgacacataaagaac 723
Qy 911 ataccaaaacagaccagctctaaaatcgatgtccacgcgtaaaagaaatgcggggctgct 970
Db 724 atcaatagcagacaccagctccgggtagatattccatagaaaagaaactctggaagtcga 783
Qy 971 gagaagtcgattactatctctactctcctgaagccactctgcgctgtgaagtcatt 1030
Db 784 gagaagcctgtcaccatccatgccccccagggggactctctgaagcatgcccgcattgatt 843
Qy 1031 ctggagattatgcataaggagctcaagatatataaaattcacagaagagatcccccttgaag 1090
Db 844 ctggaatcatgcagaagaagcagatgagaccaaactagcgaagagattctctctgaaa 903
Qy 1091 attttagctcataataactttgttgagctcttatttggtaaaagaaagaaatctttaa 1150
Db 904 atcttggcacacaatgcttgggttgaagactgattggaagaaagagcagaaattgaa 963
Qy 1151 aaaaattgacaagacacagacactaaaatcacgatatctccattgcaaggaattgagcgtg 1210
Db 964 aaaaattgaaatgaaaacagggacaaagataacaatctcatcttggaggatttgagcata 1023
Qy 1211 tataatcagaacgcactattacagtttaaagcgaatgttgagacatgtgcacagctgag 1270
Db 1024 tacaaccggaagaacacatcactgtgaagggcagagtgaggccctgtgcagtgctgag 1083
Qy 1271 gaggagatcatgaagaaatcagggagctcttatgaaatgatattgcttctatgaattctt 1330
Db 1084 atagagattatgaagagctgcgtgagcccttgagcccttgaaatgatattgctgtgtaacacc 1143
Qy 1331 caagcacatttaattcctggattaaatctgaacgccttgggtgtgttccccaccacttca 1390
```

```
Db 1144 cactc-----cggatacttctccagcctgtgtacccccatacaca 1181
Qy 1391 gggatgcacactccacactcagggcccccttccagcatgactcctccatacccgcagttt 1450
Db 1182 gttggccggtcccgcatcatcactctatc----- 1213
Qy 1451 gagcaatcagaanaacggagactgttcatcagtttatccacagctctatcagtcgggtgcac 1510
Db 1214 -----cagagcaggagattgtgaatctcttcatcccaaccaggtgtgtggcgccatc 1266
Qy 1511 atcggcaagcaggccagcacatcaacagcacttctcgttttctgaggttcattaaag 1570
Db 1267 atcgggaagaaggggcacatcaacagctggcgagattcgcgcgagcctctatcaag 1326
Qy 1571 attgctccagcgggaagcaccagatgctaaagtggagtggtgattatcactggaccacca 1630
Db 1327 atgccccctcggaagcccgagcgtcagcgaagagatggtcatcatcacccggccaccg 1386
Qy 1631 gaggtcagtttcaaggctcaggggaagaatttatggaanaatttaaagaagaaaaatttgtt 1690
Db 1387 gaagccagtttcaaggccccagggcgagctcttggaaactggaagagaggaactctttt 1446
Qy 1691 agtctctaaagaaggtgaaacttgaagctcatatcagagtgccatcttctgtcgtgac 1750
Db 1447 aaccccaagaagaagtgaaagctggaagcgcatacagagtgccccctctccacagctggc 1506
Qy 1751 agagtatttggaaaaggagggcaaacggtgaaatgaacttcagaatttgcagtgcagaa 1810
Db 1507 cggtgtgattggcaaggtggcaagccgtgaacgaactgcagaacttaaacagtgcaaa 1566
Qy 1811 gttgttgcctctcctgcacagacacctgatgagaatgacaaagtgggtgttgcataaact 1870
Db 1567 gtcatcgtgctcgtgaccacacccagatgaaatgaggaagtgcgtcagaattatc 1626
Qy 1871 ggtcacttctatgcttgcaggttggccagagaaataattcaggaattctgactcaggtta 1930
Db 1627 gggcacttcttgcagcagactgcagcgcagatcaggggaaattgtacacaggtg 1686
Qy 1931 aagcagcaccaacacagaaggctctctcaaaagtggaccac 1970
Db 1687 aagcagcaggagcagaataacctcagggagtcgctcac 1726

RESULT 5
US-09-061-709-7
; Sequence 7, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Rao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-7
```

Query Match 16.1%; Score 672.6; DB 4; Length 1946;
Best Local Similarity 66.6%; Pred. No. 9.9e-165;
Matches 979; Conservative 0; Mismatches 484; Indels 6; Gaps 1;

Qy 480 ggcgaagattcggaaacttcagatagcaaaatacccgctctattacagtggagggtgc 539
Db 103 gggttagagcggaaataatcaatccgaataatccaccacgctcgatgggaagtac 162
Qy 540 tggatattactagtcagatagagtggtgagagactgtgacaagtgaacactgact 599
Db 163 tggacagccctgctggtcagtgatggtacagtagagaactgtgagaagtgaacaccaga 222
Qy 600 cggaaactgcagttgtaaaattacactattccagtgaaggacccaagctagacaagcaactag 659
Db 223 gtgagcggcagtggtgaattgcacctattccaacccggagcagaccagccaagccatca 282
Qy 560 acaactgaatggtattccagtagagaaatttoaccttgaagtgcctatatccctgatg 719
Db 283 tgaagctgaatggccaccagttgagagaacctgccctgaaggtctctctacatccccgatg 342
Qy 720 aaatggccgcccaaaacccctctgcagcagccccagagctgcgggggcttgggcaga 779
Db 343 agcagatagcacaggccctgagaatggcgccgaggggcttggctctcctggggtcagc 402
Qy 780 ggggtcctcaaggcaggggtctccagagatccgtatccaagcagaaacctgtgatttgc 839
Db 403 ccgcagcaggtcaccctgtggcagcggggcccccagccaagcagcagcaagtggacatcc 462
Qy 840 ctctgcgcctgtggtttccaccaccaatttgttgagccatcataggaagaaaggtgcc 899
Db 463 cccttggctcctgtgtccaccaccagtatgtgggtgccattattggcaaggaggggcca 522
Qy 900 ccattcggaaacatcaacaacagaccagctcctgatacctgaagctcaagataaaattcacagaaga 1079
Db 643 gtaagatgcttggagattatgcataaagaggctgaaggacacacaaacggctgacagg 702
Qy 1080 tccccctgaagatttagctcaataaaccttgttgagcgtcttatttggttaaaagaa 1139
Db 703 tccccctgaagatcctggccataaaccttgttagggcgtctcatgtgcaaggaaagac 762
Qy 1140 gaaactttaaaaaattgacaagcacagacactaaataacagatatctctcattgcagg 1199
Db 763 ggaacctgaagaagtagacaagataccgagacaaataaccatctcctctgttgcag 822
Qy 1200 aattgacgctgtataatccagaacgcactattacagttaaaggcaatgttgagacatgtg 1259
Db 823 acctaccccttacaacctgagggaccatcactgtgaagggggcccacgagaatgtt 882
Qy 1260 ccaaaagctgaggagagatcatgaagaaatcaggagtgatttatgaaaatgattgtcctt 1319
Db 883 gcaggcccgagcaggaataatgaagaagttcggaggccctatgagaatgattgctg 942
Qy 1320 ctatgaatcttcaagcacatttaattcctgattaaatctcgaaccccttgggtctgtcc 1379
Db 943 ccatga-----gctctacactgtatccctggcctgaacctggctgtgtaggtttttcc 996
Qy 1380 caccacttcagggtgaccacctcccacctcagggccccccttcagccatgactcctccct 1439
Db 997 cagcttcatccagcgcagtcgccgcctcccagcagcgttacttgggtgctccctata 1056
Qy 1440 acccgagtttgacatcatcagaacacgagagactgttccatcagtttatcccagctctatcag 1499
Db 1057 gctcctttatgcagcgtcccgcagcagagagatggtgcagggtgtttatcccgcaggcag 1116
Qy 1500 tcggtgcatcatcggcaagcagggccagcacatcacagcagcttctcgtcttgcctgag 1559
Db 1117 tgggycgcatacatcggcaagaaggggagcacatcaaacagctctcccgcttggccagcg 1176
Qy 1560 ctccaataagattgtctccagcgggaagcaccagatgctaaagtggagtgtattatca 1619

Db 1177 cctccatcaagattgcaccaccgaaacacctgactccaaagtctogtatggttatcatca 1236
Qy 1620 ctgacacaccagaggtcagttcaaggtcaggtcaggaagatttatgaaaaattaaagaag 1679
Db 1237 ctgagaccgccagggggcccaattcaaggtcagggaaagtctatctatggaacccaaggg 1296
Qy 1680 aaaaactttgttagtctactaaagaaggtgaaacttgaagctcatatcagagtgcacatct 1739
Db 1297 agaacttcttggctcccaaggaagaagtgaaagtgcagaccacacatacgttgcgcagcat 1356
Qy 1740 ttgctgtcggcagagttattgaaaaagaggaacaaacggtaaaccttgaacttcgaatttgt 1799
Db 1357 cagcagctggccgggtcatttgcaaaagtggaacacggtgaacaggttgcagaatttga 1416
Qy 1800 caagtgcagaagttgtgtccctcgtgaccagacacctgatgagatgacaaagtgggttg 1859
Db 1417 cggcagctgaggtgtgtagcaccagagaccagaccctcgtatgagaaacagggctacog 1476
Qy 1860 tcaaaataaactggctcacttctatgcttgcagggttgcacagaaaaatttcaggaaattc 1919
Db 1477 tgaataatcatcggacatttctatgccagtcagatggtctcaacgggaatccgagacatcc 1536
Qy 1920 tgactcaggttaaaagcagcaccacaacacag 1948
Db 1537 tggccaggtttaagcagcagcatcagaag 1565

RESULT 6
US-09-061-709-5
; Sequence 5, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tsung
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; FILE REFERENCE: LUD 538
; CURRENT APPLICATION NUMBER: US/09/061.709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 5
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-5

Query Match 13.9%; Score 580.6; DB 4; Length 1708;
Best Local Similarity 66.1%; Pred. No. 6.8e-141;
Matches 857; Conservative 0; Mismatches 434; Indels 6; Gaps 1;
Qy 652 agcactgacaaaactgaatgatttcagttagagaatttcacacctgaaagtagcctatat 711
Db 37 agccatcatgaagctgaatggccaccagttgagaaacctgcccctgaaggtctctacat 96
Qy 712 cccctgatgaatggccgccagcaaaacccctctgcagcagccccgagctgcgggggct 771
Db 97 ccccgtagcagatagcagcaggggacctgagaatggggcgccgaggggcttggctctcg 156
Qy 772 tggcgagagggtcctcctaaggcaggggtctccaggatccgtatccagcagaaacacatg 831
Db 157 gggcagcccgccagggctcacctgtgtgcagcggggccccccagcagcagcagaagt 216
Qy 832 tgatttgctctgcgctgtcgtgttcccaacccaaatttggtagccatcataggaagaa 891
Db 217 ggacatcccccttggtcctcgttggtcccaacccagtatgtgggtggtccattattggcaagga 276

```
QY 892 aggtgcacacattcgaaacatcccaaaacagaccagctctataaatcgatgtccacgtaa 951
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 277 gggggccaccatccgcaatcatcaaaaacagaccagctccaaagatagacgtgcataggaa 336
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 952 agaaaatgcggggcgtcgtgagagtcgattactactactctactctactctcctgaaagccctc 1011
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 337 ggaagacgcagggtgcgcgtgcgaaagccatcagtg-gcactccaccctcgaggcgctc 396
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1012 tgcgggttaagtctctatctctgagattatgcataaggaagcctcaagatatataaaattcac 1071
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 397 ctccgcttgaagatgatcttgagattatgcataaagagcctaaagacacaaacagcgc 456
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1072 agaagatcccttgaaatttttagctcataataactttgttggacgtctctatttggttaa 1131
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 457 tgacgaggttccctgaagatcctggccataataactttgtaggcgctctcattggcaa 516
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1132 agaaggaagaatacttaaaaaattgagcaagacacagacactataaatcacgatatctcc 1191
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 517 ggaagacggaacctgaagagtagagcaagataccgagacaaaatcaccatctcctc 576
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1192 attgcggaattgacgctgtataatccagaacgcactattacagtttaaaggccaagtgtga 1251
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 577 gttgcaagaccttaccctttacaaccttgagagaccatacactgtgaaagggggccatcga 636
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1252 gacatgtccaaagctgagggagagatcatgaagaaatacaggagctcttatgaaaaatga 1311
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 637 gaattgtgcaggccgagcaggaaataatgaagaaagtcgggagccctatgagaatga 696
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1312 tattgtcttatgaattctcaagcacatttaattctcgtgattaaactctgaacgccttggg 1371
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 697 tgtggctgcata-----gctctcacgtgactccctggcctgaacctggctgctag 750
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1372 tctgttccaccacacttcaggatgccactccacctcagggcccttcagccatgac 1431
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 751 tctttccacagcttcacagcgagccgcgcgcctccagcagcgttacttggggctgc 810
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1432 tctcctcaccgcagctttagcaatcagaacggagagctgttcatcagtttaccacgc 1491
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 811 tccctatagctctctttagcagctccgcagcaggagatggcgaggtgtttatcccccgc 870
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1492 tctatcagtcggtgcacatcagcgaagcaggcggcagacacatacagcctttctcgtt 1551
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 871 ccaggcagtgggcgccatcagcgaagggcagcagacatacaacagctctcccggt 930
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1552 tgcgtgagcttcaattaaagattgctccagcggaaagcaccagatgctaaagtgagatgt 1611
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 931 tgcacgcctccataaagattgcaccacccgaaacacctgactccaaagttcgatggc 990
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1612 gattatcactgaccaccagagctcagttcaaggctcagggaagaatttatggaaaaat 1671
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 991 tatcatcactgaccgcccagagcccataatcaggctcagggaagaatctatggcaact 1050
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1672 taaagaagaaaactttgttagtctctaaagaagagtgaaactgaagctcattatcagagt 1731
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1051 caaggaggagaactcttcttggctcccaagagaggaagtgaagctggagacccacatacgt 1110
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1732 gccatcttctgctgcgcagagttatggaaaagagcgaacacggtgaatgaacttca 1791
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1111 gccagcatcagcagctgcgcgggtcattggcgaaggtggaacacggtgaaagagttgca 1170
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1792 gaattgtcaagtgcagaagttgtgtccctcgtgaccagacacccctgtatgagaatgacca 1851
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1171 gaatttgacggcagctgaggtggttagtaccagagacacagccctgtatgagaagacca 1230
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1852 agtgggtgcaaaaaacttgcgtcactctctatgcttgcaggttgcacagagaaaaattca 1911
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1231 ggtctcgtgaaaaatcctggacatttctatgccagtcagatggctcaacggaatccg 1290
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1912 ggaattctgactcaggtaaacgacgacacaaacag 1948
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1291 agaatcttggccagggttaagtcagcagcatcagaag 1327
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
```

```
RESULT 7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-Fls
US-08-232-463-14
```

```
Query Match 1.2%; Score 48.8; DB 1; Length 7218;
Best Local Similarity 4.9%; Pred. No. 0.014;
Matches 20; Conservative 217; Mismatches 169; Indels 0; Gaps 0;

QY 1661 tatggaaaattaaagaagaaactttgttagctcctaaagaagagtgaaacttgagct 1720
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1437 TACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1378
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1721 catatcagagtgccatccttctgctgctgcagagtgatttggaagaaagaggaacacggtg 1780
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1377 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1318
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1781 aatgaacttcaaatgtcgaagtgacagagttgttgcctcgtacacagacactgat 1840
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1317 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1258
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1841 gagaatgaccaaagtggttgcaaaaataactgggtaactctctatgcttgcaggttgcacag 1900
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1257 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1198
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1901 agaaaaattcagaaaattctgactcaggttaaagcagcaccacaaacagaggtctctgcaa 1960
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
```



```
Best Local Similarity 54.2%; Pred. No. 0.23; Mismatches 0; Gaps 0;
Matches 91; Conservative 0; Indels 77; Indels 0; Gaps 0;

QY 2546 tcaactacctcaggtattcagtaatacaatgaaagcaaaatgttctctttttttgaaa 2605
    || || || || || || || || || || || || || || || || || || || ||
Db 10850 TTATATATCTTAAGTTTACCTCATGTATAGTAATAAATAATTTGTCATTTTAAATATA 10791

QY 2606 attttatatactttataatgatagaagtcacacgctttttttaaataataaatt 2665
    || || || || || || || || || || || || || || || || || || || ||
Db 10790 AATTTATAATTTTAAATTTTAAAGAAAAATTTAAATTTAAAAATATAAATGATAAGA 10731

QY 2666 taacagcaatcagcaacagcgaataaagatttttacttctgctgg 2713
    || || || || || || || || || || || || || || || || || || || ||
Db 10730 AAAAGAAAAACAGTGCACCGTTGGTGTGTTTTTTTTTTTAGACACAGG 10683

RESULT 10
US-09-630-706-10/c
; Sequence 10, Application US/09630706
; Patent No. 6277640
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowert
; TITLE OF INVENTION: ANTISENSE MODULATION OF HER-3 EXPRESSION
; FILE REFERENCE: RTS-0053
; CURRENT APPLICATION NUMBER: US/09/630,706
; CURRENT FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 94
; SEQ ID NO 10
; LENGTH: 14796
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2811)...(2921)
; NAME/KEY: CDS
; LOCATION: (3174)...(3283)
; NAME/KEY: CDS
; LOCATION: (5158)...(5275)
; NAME/KEY: CDS
; LOCATION: (11955)...(12044)
US-09-630-706-10

Query Match 1.1%; Score 44.8; DB 4; Length 14796;
Best Local Similarity 54.2%; Pred. No. 0.23;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 2546 tcaactacctcaggtattcagtaatacaatgaaagcaaaatgttctctttttgaaa 2605
    || || || || || || || || || || || || || || || || || || || ||
Db 10850 TTATATATCTTAAGTTTACCTCATGTATAGTAATAAATAATTTGTCATTTTAAATATA 10791

QY 2606 attttatatactttataatgatagaagtcacacgctttttttaaataataaatt 2665
    || || || || || || || || || || || || || || || || || || || ||
Db 10790 AATTTATAATTTTAAATTTTAAAGAAAAATTTAAATTTAAAAATATAAATGATAAGA 10731

QY 2666 taacagcaatcagcaacagcgaataaagatttttacttctgctgg 2713
    || || || || || || || || || || || || || || || || || || || ||
Db 10730 AAAAGAAAAACAGTGCACCGTTGGTGTGTTTTTTTTTTTAGACACAGG 10683

RESULT 11
US-08-883-795A-36/c
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
```

```
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
US-08-883-795A-36

Query Match 1.0%; Score 43.8; DB 2; Length 665;
Best Local Similarity 45.8%; Pred. No. 0.071;
Matches 141; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 3756 tttttttcacacaatgaataaaattgctaaaaatcatggactggcttctcgttgat 3815
    || || || || || || || || || || || || || || || || || || || ||
Db 334 TTTAATTAATAAATATGTAATTTATAACATTTTAATTTATAAATATTAAATTAACAT 275

QY 3816 tcagtgatgatgtgttaaggccagagctttctcagttatgtatttttccccaata 3875
    || || || || || || || || || || || || || || || || || || || ||
Db 274 TTTAATTAATAAATATTAAATTAATAAATTTTAATTTATAAATATTAAATTAATAATAT 215

QY 3876 ttgatttttttaaaaaatatacacataggagctgcatttaaaacctgcgtgttaaaattct 3935
    || || || || || || || || || || || || || || || || || || || ||
Db 214 TTTAATTAATAAATATTAAATTAATAAATTTTAATTTATAAATATTAAATTTATAAATAT 155

QY 3936 gtcanatttcactctcagcccttttagtatggcnaaatcnaaatttactttactlaagcat 3995
    || || || || || || || || || || || || || || || || || || || ||
Db 154 TTTAATTAATAAATATTAAATTAATAAATTTTAATTTATAAATATTAAATTTATAAATAT 95

QY 3996 ttgtaattggagatctgctgactagcagaataaattcnaataattgagttttgtactc 4055
    || || || || || || || || || || || || || || || || || || || ||
Db 94 TTTAATTAATAAATATTAAATTTATAAATTTTAATTTATAAATATTAAATTTATAAATA 35

QY 4056 nccaaanatt 4064
    || ||
Db 34 TTTAATTAAT 26

RESULT 12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2002, 14:54:28 ; Search time 340.59 Seconds
(without alignments)
10524.316 Million cell updates/sec

Title: us-09-685-696-175
Perfect score: 4181
Sequence: 1 ggtgagtcgttgggtgtg.....aaaaaaaaaaaaaaaa 4181

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*

| | |
|-----|---|
| 1: | /SID88/gcgdata/geneseq/geneseq/NA1980.DAT:* |
| 2: | /SID88/gcgdata/geneseq/geneseq/NA1981.DAT:* |
| 3: | /SID88/gcgdata/geneseq/geneseq/NA1982.DAT:* |
| 4: | /SID88/gcgdata/geneseq/geneseq/NA1983.DAT:* |
| 5: | /SID88/gcgdata/geneseq/geneseq/NA1984.DAT:* |
| 6: | /SID88/gcgdata/geneseq/geneseq/NA1985.DAT:* |
| 7: | /SID88/gcgdata/geneseq/geneseq/NA1986.DAT:* |
| 8: | /SID88/gcgdata/geneseq/geneseq/NA1987.DAT:* |
| 9: | /SID88/gcgdata/geneseq/geneseq/NA1988.DAT:* |
| 10: | /SID88/gcgdata/geneseq/geneseq/NA1989.DAT:* |
| 11: | /SID88/gcgdata/geneseq/geneseq/NA1990.DAT:* |
| 12: | /SID88/gcgdata/geneseq/geneseq/NA1991.DAT:* |
| 13: | /SID88/gcgdata/geneseq/geneseq/NA1992.DAT:* |
| 14: | /SID88/gcgdata/geneseq/geneseq/NA1993.DAT:* |
| 15: | /SID88/gcgdata/geneseq/geneseq/NA1994.DAT:* |
| 16: | /SID88/gcgdata/geneseq/geneseq/NA1995.DAT:* |
| 17: | /SID88/gcgdata/geneseq/geneseq/NA1996.DAT:* |
| 18: | /SID88/gcgdata/geneseq/geneseq/NA1997.DAT:* |
| 19: | /SID88/gcgdata/geneseq/geneseq/NA1998.DAT:* |
| 20: | /SID88/gcgdata/geneseq/geneseq/NA1999.DAT:* |
| 21: | /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:* |
| 22: | /SID88/gcgdata/geneseq/geneseq/NA2001.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------------|--------------------|
| 1 | 4165 | 99.6 | 4181 | 21 AAC65900 | Human lung cancer- |
| 2 | 4143 | 99.1 | 4159 | 21 AAZ36150 | DNA encoding cance |
| 3 | 1736.8 | 41.5 | 1740 | 21 AAC66035 | Human lung cancer- |
| 4 | 804.4 | 19.2 | 2224 | 20 AAZ10617 | CDNA encoding a mu |
| 5 | 710.4 | 17.0 | 3412 | 21 AAZ36152 | DNA encoding cance |
| 6 | 695.6 | 16.6 | 3283 | 21 AAZ36154 | An alternative for |
| 7 | 672.6 | 16.1 | 1946 | 21 AAZ36153 | An alternative for |
| 8 | 580.6 | 13.9 | 1708 | 21 AAZ36151 | DNA encoding cance |
| 9 | 409.6 | 9.8 | 583 | 20 AAZ24600 | Human lung tumor a |
| 10 | 409.6 | 9.8 | 583 | 21 AAC65839 | Human lung cancer- |
| 11 | 381.4 | 9.1 | 444 | 21 AAC17226 | Human secreted pro |

| | | | | | | |
|------|-------|-----|--------|----|-----------|--------------------|
| 12 | 282 | 6.7 | 282 | 22 | AA156378 | Probe #25064 used |
| 13 | 282 | 6.7 | 588 | 22 | AA143398 | Probe #12084 used |
| 14 | 263.2 | 6.3 | 710 | 21 | AAA02565 | Human colon cancer |
| c 15 | 171.2 | 4.1 | 936 | 22 | AAF58252 | Oligonucleotide D1 |
| c 16 | 171.2 | 4.1 | 936 | 22 | AAF58254 | Oligonucleotide D1 |
| c 17 | 171.2 | 4.1 | 936 | 22 | AAF58257 | Oligonucleotide D1 |
| c 18 | 171.2 | 4.1 | 936 | 22 | AAF58259 | Oligonucleotide D2 |
| c 19 | 171.2 | 4.1 | 936 | 22 | AAF58262 | Oligonucleotide D2 |
| c 20 | 171.2 | 4.1 | 938 | 22 | AAF58255 | Oligonucleotide D1 |
| 21 | 168 | 4.0 | 936 | 22 | AAF58252 | Oligonucleotide D1 |
| 22 | 168 | 4.0 | 936 | 22 | AAF58254 | Oligonucleotide D1 |
| 23 | 168 | 4.0 | 936 | 22 | AAF58257 | Oligonucleotide D1 |
| 24 | 168 | 4.0 | 936 | 22 | AAF58259 | Oligonucleotide D2 |
| 25 | 168 | 4.0 | 936 | 22 | AAF58262 | Oligonucleotide D2 |
| 26 | 168 | 4.0 | 938 | 22 | AAF58255 | Oligonucleotide D1 |
| 27 | 163.2 | 3.9 | 364 | 21 | AAAC03267 | Human secreted pro |
| 28 | 161.2 | 3.9 | 169 | 16 | AAAT26750 | Human gene signatu |
| 29 | 137.4 | 3.3 | 500 | 22 | AA142184 | Probe #10870 used |
| 30 | 136 | 3.3 | 136 | 22 | AA155269 | Probe #23955 used |
| 31 | 121.8 | 2.9 | 424 | 21 | AAH31011 | Human colon cancer |
| 32 | 99.4 | 2.4 | 300 | 21 | AAA01526 | Human colon cancer |
| 33 | 63.4 | 1.5 | 244 | 22 | AAF58238 | Oligonucleotide D1 |
| c 34 | 63.4 | 1.5 | 244 | 22 | AAF58238 | Oligonucleotide D1 |
| c 35 | 50.8 | 1.2 | 612 | 22 | AAH71471 | Human cervical can |
| 36 | 50.6 | 1.2 | 4590 | 22 | AAH24065 | Yeast AOB9604-asso |
| 37 | 47.2 | 1.1 | 318 | 18 | AAAT91312 | Human J143-1 secre |
| 38 | 47.2 | 1.1 | 318 | 18 | AAV00436 | 3' fragment of clo |
| 39 | 46 | 1.1 | 2657 | 22 | AAH16026 | Human cDNA sequenc |
| 40 | 46 | 1.1 | 513445 | 22 | AA161373 | Soybean 318013 reg |
| 41 | 44.8 | 1.1 | 518 | 14 | AAQ48590 | Alcohol oxidase ge |
| c 42 | 44.8 | 1.1 | 14796 | 19 | AAV27941 | Survivin gene. Ho |
| 43 | 44.6 | 1.1 | 598 | 22 | AAAS05439 | Mammalian vestibul |
| c 44 | 44.2 | 1.1 | 335913 | 22 | AA161371 | Soybean 240017 reg |
| c 45 | 44.2 | 1.1 | 335913 | 22 | AA161372 | Soybean 240017 reg |

ALIGNMENTS

RESULT 1

AAC65900

ID AAC65900 standard; cDNA; 4181 BP.

XX

AC AAC65900;

XX

XX

DT 21-FEB-2001 (first entry)

XX

DE Human lung cancer-associated cDNA L523S.

XX

KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;

KW vaccine; detection; ss.

XX

OS Homo sapiens.

XX

PN WO200061612-A2.

XX

XX

PD 19-OCT-2000.

XX

PF 03-APR-2000; 2000WO-US08896.

XX

PR 02-APR-1999; 99US-0285479.

PR 17-DEC-1999; 99US-0466396.

PR 30-DEC-1999; 99US-0476496.

PR 10-JAN-2000; 2000US-0480884.

PR 22-FEB-2000; 2000US-0510376.

XX

(CORI-) CORIXA CORP.

PI Wang T, Fan L;

XX

DR WPI; 2000-628399/60.

DR P-PSDB; AAB11328.

XX

Db 1801 |||||aagtcagaaagttgttgccttcgtgaccagacacatgtagaatgaccaagtggtgt 1860
 Qy 1861 caaaaataactggtccactctatgcttgccaggttgcacagaaaaattcaggaaattct 1920
 Db 1861 caaaaataactggtccactctatgcttgccaggttgcacagaaaaattcaggaaattct 1920
 Qy 1921 gactcaggtaagcagcaccacaacagaaaggtctctgcaaaagtgacacacctcagtcag 1980
 Db 1921 gactcaggtaagcagcaccacaacagaaaggtctctgcaaaagtgacacacctcagtcag 1980
 Qy 1981 acggaagttaagcgtcaggaaacagccacacagagcagagtgccaaacccaagaacaga 2040
 Db 1981 acggaagttaagcgtcaggaaacagccacacagagcagagtgccaaacccaagaacaga 2040
 Qy 2041 ttgcttaacacaacagatggcgctgaccccttatccagaatcacatgcacaagtttttac 2100
 Db 2041 ttgcttaacacaacagatggcgctgaccccttatccagaatcacatgcacaagtttttac 2100
 Qy 2101 ctagccagttgttctgaggaacaggcaacttttgaactcctgtctctgtggaattgtat 2160
 Db 2101 ctagccagttgttctgaggaacaggcaacttttgaactcctgtctctgtggaattgtat 2160
 Qy 2161 actttatgctctgaaatgtagacacccagcttttaaaacaaaacaaacaaacaaa 2220
 Db 2161 actttatgctctgaaatgtagacacccagcttttaaaacaaaacaaacaaacaaa 2220
 Qy 2221 aaaaaggtgggaggggaaagagaagagctctgcacttccctttgttagtctca 2280
 Db 2221 aaaaaggtgggaggggaaagagaagagctctgcacttccctttgttagtctca 2280
 Qy 2281 cagtaataacagatatcttaattcttcttaattcccccataatgccagaaattggtctta 2340
 Db 2281 cagtaataacagatatcttaattcttcttaattcccccataatgccagaaattggtctta 2340
 Qy 2341 atgagtcttctactaaatctcaaaatagatgctccttaaatccaaattgttaaaattgga 2400
 Db 2341 atgagtcttctactaaatctcaaaatagatgctccttaaatccaaattgttaaaattgga 2400
 Qy 2401 tcagataaatatcacaggaacttaaatgttaagccattagcatagaaaaacttgttctca 2460
 Db 2401 tcagataaatatcacaggaacttaaatgttaagccattagcatagaaaaacttgttctca 2460
 Qy 2461 gttttatttttacctaaactaacatgagtaacctaaagggaagtgctgaattggttggc 2520
 Db 2461 gttttatttttacctaaactaacatgagtaacctaaagggaagtgctgaattggttggc 2520
 Qy 2521 aggggtattaaacgtgcatttttactccaactcactcaggtattcagtaatacaatgaaa 2580
 Db 2521 aggggtattaaacgtgcatttttactccaactcactcaggtattcagtaatacaatgaaa 2580
 Qy 2581 gcaaaaattgttcctttttttgaaaaatttatatacttttataatgtagaagtcacaacg. 2640
 Db 2581 gcaaaaattgttcctttttttgaaaaatttatatacttttataatgtagaagtcacaacg 2640
 Qy 2641 ttttttaaaaaataaatttaaaatttaacagcaatcagctaacagggcaaatlaagatttt 2700
 Db 2641 ttttttaaaaaataaatttaaaatttaacagcaatcagctaacagggcaaatlaagatttt 2700
 Qy 2701 tactcttgctgggtgacagtaaaagctggaaaaatttaatttcagggttttttgagggtttcg 2760
 Db 2701 tactcttgctgggtgacagtaaaagctggaaaaatttaatttcagggttttttgagggtttcg 2760
 Qy 2761 acacagttattagttaaatcaaatgttcaaaaaatacaggagcagtgctagatcttgaga 2820
 Db 2761 acacagttattagttaaatcaaatgttcaaaaaatacaggagcagtgctagatcttgaga 2820
 Qy 2821 gcagcactaccatttattctttcatttatagttgggaaagttttttgacggtactaaca 2880
 Db 2821 gcagcactaccatttattctttcatttatagttgggaaagttttttgacggtactaaca 2880
 Qy 2881 gtggtcgaggaagatttttggaaacgctggttttaaatggcttcagagagactcagttttt 2940
 Db |||||

Db 2881 gtggtcgaggagatttttggaaacgctggttttaattggcttcaggagactcagttttt 2940
 Qy 2941 gttagctacatgattgaatgataataaagcttctgtctctgactatcaatacctaa 3000
 Db 2941 gttagctacatgattgaatgataataaagcttctgtctctgactatcaatacctaa 3000
 Qy 3001 agaaaatgcacagtcagtgaaagatgcaagacttttcaactgactggcaaaaagcaagcttta 3060
 Db 3001 agaaaatgcacagtgaaagatgcaagacttttcaactgactggcaaaaagcaagcttta 3060
 Qy 3061 gcttgccttataggatgcttagtttgcactacacatccagaccaaatgggacagtcacataga 3120
 Db 3061 gcttgccttataggatgcttagtttgcactacacatccagaccaaatgggacagtcacataga 3120
 Qy 3121 tgggtgacagtgcttttaaacgcaaaaaggtcacatttccatggggccagcactgtcat 3180
 Db 3121 tgggtgacagtgcttttaaacgcaaaaaggtcacatttccatggggccagcactgtcat 3180
 Qy 3181 gagcctcactaagctattttgaagatttttaagcactgataaattaaaaaataaaaaaa 3240
 Db 3181 gagcctcactaagctattttgaagatttttaagcactgataaattaaaaaataaaaaaa 3240
 Qy 3241 aaattagactccaccttaagtagtaaaagtataacagagatttctgtactgtgcaatcag 3300
 Db 3241 aaattagactccaccttaagtagtaaaagtataacagagatttctgtactgtgcaatcag 3300
 Qy 3301 ttctttgaaaaaaaagtcaaaagatagagaatacaagaaaaagttttngggataataattg 3360
 Db 3301 ttctttgaaaaaaaagtcaaaagatagagaatacaagaaaaagttttngggataataattg 3360
 Qy 3361 aatgactgtgaaaaacatatgacctttgataaacgaactcatttgcctcactccttgacagca 3420
 Db 3361 aatgactgtgaaaaacatatgacctttgataaacgaactcatttgcctcactccttgacagca 3420
 Qy 3421 aagccacgtacgtacaatgtgtgtgggtgtgggtgtgtcccaagggccagcgtctctctg 3480
 Db 3421 aagccacgtacgtacaatgtgtgtgggtgtgggtgtgtcccaagggccagcgtctctctg 3480
 Qy 3481 aattgatttttgagtttttggnttgnaaagtatcacacagncatgttacactgactctnaa 3540
 Db 3481 aattgatttttgagtttttggnttgnaaagtatcacacagncatgttacactgactctnaa 3540
 Qy 3541 ggacatatntataaacaccttttaaaaaaaatccccctgctcattcttatttcgagatga 3600
 Db 3541 ggacatatntataaacaccttttaaaaaaaatccccctgctcattcttatttcgagatga 3600
 Qy 3601 atttcgatacagactagatgtcttctgaaagatcaattagacattntgaaaaatgatttaa 3660
 Db 3601 atttcgatacagactagatgtcttctgaaagatcaattagacattntgaaaaatgatttaa 3660
 Qy 3661 agtggttttcccttaagtctctgaaaaacagtttctttttagtttttaacccaaaaagtg 3720
 Db 3661 agtggttttcccttaagtctctgaaaaacagtttctttttagtttttaacccaaaaagtg 3720
 Qy 3721 ccccttttgcactgggttctcctagcattcatgatttttttttcacacaaatgaattaaa 3780
 Db 3721 ccccttttgcactgggttctcctagcattcatgatttttttttcacacaaatgaattaaa 3780
 Qy 3781 attgctaaaaatcatggactggcgttctctggtgatttcaggtgaagtgtgtttaaggcca 3840
 Db 3781 attgctaaaaatcatggactggcgttctctggtgatttcaggtgaagtgtgtttaaggcca 3840
 Qy 3841 gagcctttctcagttatttgaattttttccccaatttttgaatttttlaaaaatacacat 3900
 Db 3841 gagcctttctcagttatttgaattttttccccaatttttgaatttttlaaaaatacacat 3900
 Qy 3901 aggagctgcatatttaaaacctgctgggttttaattctctcanatttccactctagcctttta 3960
 Db 3901 aggagctgcatatttaaaacctgctgggttttaattctctcanatttccactctagcctttta 3960
 Qy 3961 gtatgcnatcanaanatttacttttacttaagcatttggagatttggagattctgggtacta 4020
 Db gtatgcnatcanaanatttacttttacttaagcatttggagatttggagattctgggtacta 4020

```
QY 4021 gctaagaaataattcnataattgagttgtactcncncaaanatggtcattctctcatgn 4080
DB 4021 gctaagaaataattcnataattgagttgtactcncncaaanatggtcattctctcatgn 4080
QY 4081 ataagtgtcncncccaatcagcttctatttccagacnaccttgacgcagataaaatttttc 4140
DB 4081 ataagtgtcncncccaatcagcttctatttccagacnaccttgacgcagataaaatttttc 4140
QY 4141 atcatttaggtcccccacaaaaaaaccccccccccccccccccccccccccccccc 4181
DB 4141 atcatttaggtcccccacaaaaaaaccccccccccccccccccccccccccccccc 4181

RESULT 2
ID AA236150
XX AA236150 standard; DNA; 4159 BP.
AC AA236150;
XX
DT 11-FEB-2000 (first entry)
XX
DE DNA encoding cancer associated antigen KOC-1.
XX
KW Cancer associated antigen; KOC-1; cancer; vaccine; CT7; ss.
XX
OS Homo sapiens.
XX
PN WO9954738-A1.
XX
PD 28-OCT-1999.
XX
PF 16-MAR-1999; 99WO-US05766.
XX
PR 17-APR-1998; 98US-0061709.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX WPI; 2000-013284/01.
XX
PT Nucleotides representing cancer-associated genes, used to develop
PT products for the diagnosis, monitoring and treatment of cancers -
XX
PS Claim 88; Page 39-40; 44pp; English.
XX
CC The present sequence represents a cancer associated antigen gene
CC designated KOC-1. The specification also describes a cancer associated
CC antigen designated CT7. The CT7 polynucleotide was isolated from
CC SK-MEL-37 melanoma cells. The polypeptide has some homology with
CC MAGE-10, limited to about 210 carboxy terminal amino acids. The amino
CC terminal of the protein has a repetitive pattern, with repeats rich in
CC serine, proline, glutamine and leucine, and an almost invariable core of
CC the peptide given in AY43877. The CT7 polypeptide can be processed to
CC peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC and polypeptides can be used for treating a cancerous condition and
CC screening for or diagnosing cancerous conditions. The cancer associated
CC antigens can be used as an immunogenic or vaccine composition with an
CC adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
CC stimulating factor (GM-CSF).
XX
SQ Sequence 4159 BP; 1281 A; 830 C; 851 G; 1181 T; 16 other;
```

```
Query Match 99.1%; Score 4143; DB 21; Length 4159;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 4159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ggtggatgcgttgggttagctagctttttctttttctttttcttttttaaacacatct 60
DB 1 ggtggatgcgttgggttagctagctttttctttttctttttcttttttaaacacatct 60
```

```
QY 61 agacaagaaaaaacagcctcggtatctgatttttctactcctcgtcttcttgcttggttc 120
DB 61 agacaagaaaaaacagcctcggtatctgatttttctactcctcgtcttcttgcttggttc 120
QY 121 ttactgtttgtgtattttaaagcgagaagcagagggggaacaaaacacagctgagatcca 180
DB 121 ttactgtttgtgtattttaaagcgagaagcagagggggaacaaaacacagctgagatcca 180
QY 181 tccatcacctgggtgttttaatttttctgtttttctcgttatttttttaacaacc 240
DB 181 tccatcacctgggtgttttaatttttctgtttttctcgttatttttttaacaacc 240
QY 241 actcttcacaatgaacaaactgtatatcggaaacctcagcgagaaacgcgccccctcgga 300
DB 241 actcttcacaatgaacaaactgtatatcggaaacctcagcgagaaacgcgccccctcgga 300
QY 301 cctagaagtatcttcaaggaacccaagatcccggtgtcgggaccccttctcgttgaagac 360
DB 301 cctagaagtatcttcaaggaacccaagatcccggtgtcgggaccccttctcgttgaagac 360
QY 361 tggctacgcgttcgtgactgcgcggaagctgggcccctcaaggccatcagaggcgt 420
DB 361 tggctacgcgttcgtgactgcgcggaagctgggcccctcaaggccatcagaggcgt 420
QY 421 ttcaggtaaaatagaactgcacgggaaacccatagaagttagcactcgggtcccaaaaag 480
DB 421 ttcaggtaaaatagaactgcacgggaaacccatagaagttagcactcgggtcccaaaaag 480
QY 481 gcaaggaattcggaaacttcagatacgaataatccccccctatttacagtggggaggtc 540
DB 481 gcaaggaattcggaaacttcagatacgaataatccccccctatttacagtggggaggtc 540
QY 541 ggtagtttactagtcagtgatggagtggtagagctgtgagcaagtgaacactgactc 600
DB 541 ggtagtttactagtcagtgatggagtggtagagctgtgagcaagtgaacactgactc 600
QY 601 ggaactgcagtggttaaatgaactattccagtaaggaaacacagctagacaagcactaga 660
DB 601 ggaactgcagtggttaaatgaactattccagtaaggaaacacagctagacaagcactaga 660
QY 661 caaactgaatggatttcagttagagaatttcaccttgaagctagcctatccccctgata 720
DB 661 caaactgaatggatttcagttagagaatttcaccttgaagctagcctatccccctgata 720
QY 721 aatggccgccagcaaaccccttgccagcagcccgaggctgcgcggggttggcgagag 780
DB 721 aatggccgccagcaaaccccttgccagcagcccgaggctgcgcggggttggcgagag 780
QY 781 gggctctcaaggcaggggtctccagatccgtatccaaagcagaaacacatgtgattgcc 840
DB 781 gggctctcaaggcaggggtctccagatccgtatccaaagcagaaacacatgtgattgcc 840
QY 841 tctgcgcctcgtgttcccccacaaatttctgagccatcataggaagaaagtgccac 900
DB 841 tctgcgcctcgtgttcccccacaaatttctgagccatcataggaagaaagtgccac 900
QY 901 cattcggaacatcaccaaaaacagaccctgtctaaatcgatgtccaccgttaaaagaaatgc 960
DB 901 cattcggaacatcaccaaaaacagaccctgtctaaatcgatgtccaccgttaaaagaaatgc 960
QY 961 gggggtcgtgagaagtcgattactatccctctactcctcctgaaggacccctctcggttg 1020
DB 961 gggggtcgtgagaagtcgattactatccctctactcctcctgaaggacccctctcggttg 1020
QY 1021 taagtctattctggagattatgcataaggaagctcaagataataaaatccacagagagat 1080
DB 1021 taagtctattctggagattatgcataaggaagctcaagataataaaatccacagagagat 1080
QY 1081 ccccttgaagatttttagctcataataaactttgttggagctcttatttggtaaaagaaag 1140
DB 1081 ccccttgaagatttttagctcataataaactttgttggagctcttatttggtaaaagaaag 1140
QY 1141 aatctttaaaaaaattgagcaagacacagacactaaataatcacgatatctcctccatgcagga 1200
```


Db 1141 aatcttaaaaaattgagcacagacacactaaaaatcacgatatctccattgcagga 1200
Qy 1201 attgacgtgtataatccagaaacgacactattacaggttaaaggcaatgttgagacatgtgc 1260
Db 1201 attgacgtgtataatccagaaacgacactattacaggttaaaggcaatgttgagacatgtgc 1260
Qy 1261 caaagctgaggagatcatgaagaaatcaggagctcttatgaaaaatgatattgcttc 1320
Db 1261 caaagctgaggagatcatgaagaaatcaggagctcttatgaaaaatgatattgcttc 1320
Qy 1321 tatgaattctcaagcacatttaattccctggattaaatctgaacgcttgggtctgttccc 1380
Db 1321 tatgaattctcaagcacatttaattccctggattaaatctgaacgcttgggtctgttccc 1380
Qy 1381 acccaactcaaggatgcacactcccacactcagggccttcagccatgcactccctcta 1440
Db 1381 acccaactcaaggatgcacactcccacactcagggccttcagccatgcactccctcta 1440
Qy 1441 cccgcagtttgagcaatcagaacaggagactgttcatcagtttatcccagctctatcagt 1500
Db 1441 cccgcagtttgagcaatcagaacaggagactgttcatcagtttatcccagctctatcagt 1500
Qy 1501 cagtgccatcatcggcaagcaggccagcacatcaagcagcttctcgtcttgctggagc 1560
Db 1501 cagtgccatcatcggcaagcaggccagcacatcaagcagcttctcgtcttgctggagc 1560
Qy 1561 tccaattaaagtgtctcagcggaagcaccagatgctaaagtggagatgggtgattatcac 1620
Db 1561 tccaattaaagtgtctcagcggaagcaccagatgctaaagtggagatgggtgattatcac 1620
Qy 1621 tggaccaccagaggtcagttcaaggctcagggaagaatttatgaaaaatcaagaaga 1680
Db 1621 tggaccaccagaggtcagttcaaggctcagggaagaatttatgaaaaatcaagaaga 1680
Qy 1681 aaactttgttagcttaagaagaaggtgaaactgaaagctcaatcaagtgccatcctt 1740
Db 1681 aaactttgttagcttaagaagaaggtgaaactgaaagctcaatcaagtgccatcctt 1740
Qy 1741 tgcctgtgcagagttatttgaaaaggaggcaaaacgttgaaatgaactcagaatttgc 1800
Db 1741 tgcctgtgcagagttatttgaaaaggaggcaaaacgttgaaatgaactcagaatttgc 1800
Qy 1801 aagtcagaaagtgtgtccctcgtgacacagacactgatgagaatgaccaagtgtgtg 1860
Db 1801 aagtcagaaagtgtgtccctcgtgacacagacactgatgagaatgaccaagtgtgtg 1860
Qy 1861 caaataactgctactctetagtctgcoaggtgtcccagagaaaaatcaggaattct 1920
Db 1861 caaataactgctactctetagtctgcoaggtgtcccagagaaaaatcaggaattct 1920
Qy 1921 gactcaggttaaagcagcaccacaacagaggctctgcaaaagtgaacctcagtcacag 1980
Db 1921 gactcaggttaaagcagcaccacaacagaggctctgcaaaagtgaacctcagtcacag 1980
Qy 1981 acggaagtaaaggtcaggaaacagccaccacagaggcagatgccaaaccaaaagacaga 2040
Db 1981 acggaagtaaaggtcaggaaacagccaccacagaggcagatgccaaaccaaaagacaga 2040
Qy 2041 ttgcttaaccaacagatgggogctgacccccctatccagaatcacatgcacaagttttac 2100
Db 2041 ttgcttaaccaacagatgggogctgacccccctatccagaatcacatgcacaagttttac 2100
Qy 2101 ctagccagttgttctgaggaccaggaacttttgaactcctgtctctgtgagaatgtat 2160
Db 2101 ctagccagttgttctgaggaccaggaacttttgaactcctgtctctgtgagaatgtat 2160
Qy 2161 actttatgctctgaaatgtatgacaccagctttaaacacaaacaaacaaacacaaa 2220
Db 2161 actttatgctctgaaatgtatgacaccagctttaaacacaaacaaacacaaacaaa 2220
Qy 2221 aaaaaggtggggagggaagagagagctctgcacttccctttgtgtagtctca 2280

Db 2221 aaaaaggtggggagggaagagagagctctgcacttccctttgtgtagtctca 2280
Qy 2281 caagtaaacagatatcttaattctcttaattatcccccataatgcagaaattggctta 2340
Db 2281 caagtaaacagatatcttaattctcttaattatcccccataatgcagaaattggctta 2340
Qy 2341 atgagtctttcactaaattcatcaaatagattgctcctaaaaatcccaattgttaaaaattgga 2400
Db 2341 atgagtctttcactaaattcatcaaatagattgctcctaaaaatcccaattgttaaaaattgga 2400
Qy 2401 tcagaataattatcacaggaacttaattgtaagccattagcatagaaaaactgttctca 2460
Db 2401 tcagaataattatcacaggaacttaattgtaagccattagcatagaaaaactgttctca 2460
Qy 2461 gttttatttttaccataaactaacatgagtaacctaaagggaagtgctgaatgggtgtgc 2520
Db 2461 gttttatttttaccataaactaacatgagtaacctaaagggaagtgctgaatgggtgtgc 2520
Qy 2521 aggggtattaaacgtgcatttttactcaactcctcaggtattcagtaatacaatgaaa 2580
Db 2521 aggggtattaaacgtgcatttttactcaactcctcaggtattcagtaatacaatgaaa 2580
Qy 2581 gaaaaattgttccctttttttgaaaaatttatactttataatgataagtcacaacg 2640
Db 2581 gaaaaattgttccctttttttgaaaaatttatactttataatgataagtcacaacg 2640
Qy 2641 ttttttaaaaaataaatttaaaatttaacagcaatcaagcaatacaggaataaagatttt 2700
Db 2641 ttttttaaaaaataaatttaaaatttaacagcaatcaagcaatacaggaataaagatttt 2700
Qy 2701 tacttctggctggtagacagtaagctggaaaatttaatttcagggttttttgaggctttg 2760
Db 2701 tacttctggctggtagacagtaagctggaaaatttaatttcagggttttttgaggctttg 2760
Qy 2761 acacagttattagttaaatcacaatgttcaaaaaatacaggagcagtgctagatctggaga 2820
Db 2761 acacagttattagttaaatcacaatgttcaaaaaatacaggagcagtgctagatctggaga 2820
Qy 2821 gcagcactaccatttattcttcaatttagttgggaaagtttttgacggtaactaaca 2880
Db 2821 gcagcactaccatttattcttcaatttagttgggaaagtttttgacggtaactaaca 2880
Qy 2881 gtggtcgcaggagattttggaaacgctgtttaaattggcttcaggagactcagttttt 2940
Db 2881 gtggtcgcaggagattttggaaacgctgtttaaattggcttcaggagactcagttttt 2940
Qy 2941 gtttgctacatgattgaaatgataataaaatgctttgtctctgactatcaataactaa 3000
Db 2941 gtttgctacatgattgaaatgataataaaatgctttgtctctgactatcaataactaa 3000
Qy 3001 agaaagtcacagtggaagatgcaagacttccaactgactggcaaaagcaagcttta 3060
Db 3001 agaaagtcacagtggaagatgcaagacttccaactgactggcaaaagcaagcttta 3060
Qy 3061 gcttgcttattaggtgcttagttggccactacacttcagaccaatgggacagctcata 3120
Db 3061 gcttgcttattaggtgcttagttggccactacacttcagaccaatgggacagctcata 3120
Qy 3121 tgggtgacagtggtttaaacycaaaaaaggctacatttccatggggccagcactgcat 3180
Db 3121 tgggtgacagtggtttaaacycaaaaaaggctacatttccatggggccagcactgcat 3180
Qy 3181 gggcctcactaagctattttgaagatttttaagcactgataaataaaaaaataaaaaa 3240
Db 3181 gggcctcactaagctattttgaagatttttaagcactgataaataaaaaaataaaaaa 3240
Qy 3241 aaattagactccacttaagttagtaaaagtataacaggtttctgtactgtgcaatcag 3300
Db 3241 aaattagactccacttaagttagtaaaagtataacaggtttctgtactgtgcaatcag 3300
Qy 3301 ttctttgaaaaaaagctcaaaagatagagaaatacaagaaaagttttngggatataattg 3360
Db 3301 ttctttgaaaaaaagctcaaaagatagagaaatacaagaaaagttttngggatataattg 3360

| | | | | |
|---|---|------|---|------|
| D | b | 1501 | agagtattgaaaggaggcaaacacgggtgaatgaacttcagaataattgtcgaatgcagaa | 1560 |
| O | y | 1811 | gttgtgtccctcgtagaccagacacctgatgagaatgaccaagtgttgcataataaact | 1870 |
| D | b | 1561 | gttgtgtccctcgtagaccagacacctgatgagaatgaccaagtgttgcataataaact | 1620 |
| O | y | 1871 | ggtaacttttatcttgccaggtgtcccagagaaaaaatcaggaaaattctgactcaggtta | 1930 |
| D | b | 1621 | ggtaacttttatcttgccaggtgtcccagagaaaaaatcaggaaaattctgactcaggtta | 1680 |
| O | y | 1931 | aagcagcaccaacaacagaaagctctgcaaatggacacacctcagtcgaagcgaagtaa | 1990 |
| D | b | 1681 | aagcagcaccaacaacagaaagctctgcaaatggacacacctcagtcgaagcgaagtaa | 1740 |
| RESULT 4 AAZ10617 ID AAZ10617 standard; cDNA; 2224 BP. XX AC AAZ10617; XX AC XX XX DT 17-NOV-1999 (first entry) XX XX cDNA encoding a murine c-myc coding region determinant binding protein DE XX c-myc coding region determinant binding protein; CRD-BP; tumor; KW KW c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer; KW KW pancreatic cancer; ss. XX OS Mus musculus. XX FH Key Location/Qualifiers FT CDS 131..1864 ET /tag= a XX PN W09946594-A2. XX PD 16-SEP-1999. XX PF 05-MAR-1999; 99WO-US04897. XX PR 09-MAR-1998; 98US-0077372. XX PA (WISC) WISCONSIN ALUMNI RES FOUND. XX PI Ross J; DR WPI; 1999-551506/46. DR P-PSDB; AAY30649. XX PT Diagnosing presence or absence of a tumor in a human by examining c-myc PT coding region determinant-binding protein - XX PS Example; Fig 1A-D; 79pp; English. XX CC The present sequence encodes a murine c-myc coding region determinant CC binding protein (CRD-BP). The presence or absence of a tumor can be CC determined by determining the levels of CRD-BP present in the suspect CC tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack CC and so prolongs its half-life. The methods are used for diagnosing CC presence or absence of a tumor in a human, especially breast, colon CC and pancreatic cancer. They are also used to inhibit cancer cell CC growth. | | | | |
| X | S | Q | Sequence 2224 BP; 580 A; 609 C; 618 G; 417 T; 0 other; | |
| Query Match 19.2%; Score 804.4; DB 20; Length 2224; Best Local Similarity 67.0%; Pred. No. 1.2e-170; Matches 1141; Conservative 0; Mismatches 561; Indels 0; Gaps | | | | |
| O | y | 247 | cacaatgaacaaactgtatatcggaacctcagcgagaacgccgccctcgaccta | 306 |

Db 127 caccatgaacaagctttacatcgcgcgaacctcaacgagagtgtgacccccgcagacttggga 186
Qy 307 aagtattctcaagcagcccaagatccccggtgtcgggacctctcctgttgaaagactggcta 366
Db 187 gaaagtattcgcggagcacagaatctctacagcgcgcagttcttggtcacaaatccggcta 246
Qy 367 cgcgttcgtgacttcgcggagcagagactgggcccctcaaggccactcagggcgttttcagg 426
Db 247 cgccttcgtggattgcgccacagcagctggcgatggaaggccatcgaaaactttctcggg 306
Qy 427 taaaatagaactgcacgggaaacccaTagaagttgagcactcggctcccaaaaggcaag 486
Db 307 gaaagtagaactgcaaggaaaacgcttagagattgaaacactcagtcctcccaaaaacaag 366
Qy 487 gattcggaaacttcagatacgaataatccccgcctcatttacagtgggaggtgctgggtag 546
Db 367 gagtgcgaaaatacagatccgcgaataatccacctcagctccgagtgggaggtgctagatag 426
Qy 547 ttactagtccagtatagagtgtggagagctgtgagcaagtgagcaactgactcgggaaac 606
Db 427 cctgctggctcagtcacgttacagtgggaaactgtgacaagtgaaacactgaaagtggac 486
Qy 607 tgcagttgtaaatgaacctattccagtaagaccagagctagacaagcacttagacaact 666
Db 487 agcgtgtcaactcactactactaaccgaggagcagaccaggcaagctatcatgaagct 546
Qy 667 gaaatggtattcagtagagatttcaaccttgaaagttagcctatatcctcgtatgaatggc 726
Db 547 aaatggccataactggagaacctgcccgaaggtctcctacatacctgatgagcagat 606
Qy 727 cgcgcagcaaaaccttgacagccccgaggtcgcgggggcttgggcagaggggctc 786
Db 607 aacacaggttcctgagaaatggcgctgtgaggtcttgggtcgtgggcccagcccggca 666
Qy 787 ctcaaggcagggttctccagagctcgtatccaaagcagaacactgtgatttgcctctgcg 846
Db 667 aggtgcgccgtggcagcaggggtccagccaagcagcagccagtgacatccctctccg 726
Qy 847 cctgctggttcccaacaaattgttgagccatcatataggaaaagaggtgccaccattcg 906
Db 727 gctcctggtgctacgcagtatgtaggcgtatcatcttggcaaggaggtgccaccatccg 786
Qy 907 gaacatcaccaaacagaccagctcaaaatcgatgtccaccgttaaaagaaatcgggggc 966
Db 787 aaacatcacaaaacagcagctccaaaatagacgtgcatagaaagagaaatcgggcgc 846
Qy 967 tgcgtgaagtcgattactactcctctactcctgaaggcacctctcgcgttgaagtc 1026
Db 847 tgcgagaagccatcagctgcatccaacctgaaggctcctcctcgcgtgcaagat 906
Qy 1027 tattctggagattatgcataagaagctcaagataataaaattcacagaagatccccctt 1086
Db 907 gatctggagattatgcacaaggaggcaagagacacacaaacggcagatgaagttccct 966
Qy 1087 gaagattttagctcaataaactttgttgacgtcttatttggttaaaagaagaagaaatct 1146
Db 967 gaagatcctggtcctaacaactctcgtcggcgactcatttggcaagaaaggccggaacct 1026
Qy 1147 taaaaaatlgacaagacacacacataaaatcacgatatctcctcattgacagaattgac 1206
Db 1027 gaagaaggtggagcagacagagacgagatccacatctcctcctcagaccctcac 1086
Qy 1207 gctgtataatccgaacgcactattacagttaaaggcaatgttgagacatgtgccaaagc 1266
Db 1087 gctctaaacctgagaggaccatcactgtgaaggcgccatgagaactgttgagggc 1146
Qy 1267 tgaggaggagatcatgaagaaaatcaggaggtcttatgaaaaatgatattgtctctatgaa 1326
Db 1147 cgaagcaggagatcatgaagaaagtctgagaggttactcagaacgacgtggccgcctag 1206
Qy 1327 tcttcagcacatttaattcctgattaaatctgaacgcttggtctgtctccaccac 1386
Db 1207 cttgcggtcccaacctatccctgggttaaccttgcgtgaggtctctctccagcttc 1266

RESULT 5

AAZ36152

ID AAZ36152 standard; DNA; 3412 BP.

XX

AC AAZ36152;

XX

DT 11-FEB-2000 (first entry)

XX

DE DNA encoding cancer associated antigen KOC-3.

XX

KW Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.

XX

OS Homo sapiens.

XX

PN WO9954738-A1.

XX

PD 28-OCT-1999.

XX

PF 16-MAR-1999; 99WO-US05766.

XX

PR 17-APR-1998; 98US-0061709.

XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

XX

DR WPI; 2000-013284/01.

XX

PT Nucleotides representing cancer-associated genes, used to develop

PT products for the diagnosis, monitoring and treatment of cancers -

XX

Db 1057 gctcctttatgcaggtcccgagcagagagatggtgcagggtgtttatcccccagcag 1116
Qy 1500 tccgtgcacatcggcgaagcagggccgacacatcacagcagcttctcgtcttgcaggag 1559
Db 1117 tggcgcccatcgcgcaagagggcgacacatcacacagcctcccggttgcagcg 1176
Qy 1560 cttaataatgattgctccagcggaaacaccagatgctaaagtgcagatggtgattatca 1619
Db 1177 cctccataagattgcaaccaccgcaaacactgactccaagtctcgtatggttatcatca 1236
Qy 1620 ctggaccaccagaggtcagctcaaggctcagggaacatttatggaaaaattaaagaag 1679
Db 1237 ctgacgcgcagagggcccaattcaaggctcagggaagaattctatggcaaacatcaaggag 1296
Qy 1680 aaaactttgttagcttaagaagaggtgaaacttgaagctcatatcagagtgccatcct 1739
Db 1297 agaactcttttggcccaaggaggaagtgaagctggagaccacatcagtgccagcat 1356
Qy 1740 tgcgtctgcagagttattggaaaaagggcaaaacgggtgaatgaacttcagaatttgt 1799
Db 1357 cagcagctggccgggtcattggcaaaagtggaacacgggtgaacgagttgcagaatttga 1416
Qy 1800 caagtgcagaagttgttccctctgaccagacacactgatgagaatgaccaagtgtgtg 1859
Db 1417 cggcagctgaggtgtagtaccaaagagaccagaccctctgatggaaagaccaggtcatcg 1476
Qy 1860 tcaaaataactggtcactcttatctgttccaggttgcagagagaaaaattcaggaaattc 1919
Db 1477 tgaataatcatcgacattctatccagtcagatggctcaacggaagatccgagacatcc 1536
Qy 1920 tgactcaggttaagcagcaccacaacacag 1948
Db 1537 tggcccaaggttaagcagcagcatcagaag 1565

RESULT 8

AAZ36151

ID AAZ36151 standard; DNA; 1708 BP.

AC AAZ36151;

XX 11-FEB-2000 (first entry)

DE DNA encoding cancer associated antigen KOC-2.

XX Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.

XX Homo sapiens.

XX W09954738-A1.

XX 28-OCT-1999.

XX 16-MAR-1999; 99WO-US05766.

XX 17-APR-1998; 98US-0061709.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

XX WPI; 2000-013284/01.

XX Nucleotides representing cancer-associated genes, used to develop
PT products for the diagnosis, monitoring and treatment of cancers

XX Claim 55; Page 40; 44pp; English.

XX The present sequence represents a cancer associated antigen gene
CC designated KOC-2. The specification also describes a cancer associated
CC antigen designated CT7. The CT7 polynucleotide was isolated from
CC SK-MEL-37 melanoma cells. The polypeptide has some homology with
CC MAGE-10, limited to about 210 carboxy terminal amino acids. The amino

CC terminal of the protein has a repetitive pattern, with repeats rich in
CC serine, proline, glutamine and leucine, and an almost invariable core of
CC the peptide given in AAY43877. The CT7 polypeptide can be processed to
CC peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC and polypeptides can be used for treating a cancerous condition and
CC screening for or diagnosing cancerous conditions. The cancer associated
CC antigens can be used as an immunogenic or vaccine composition with an
CC adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
CC stimulating factor (GM-CSF).
XX
SQ Sequence 1708 BP; 447 A; 469 C; 473 G; 314 T; 5 other;

Query Match 13.9%; Score 580.6; DB 21; Length 1708;

Best Local Similarity 66.1%; Pred. No. 1.7e-120;

Matches 857; Conservative 0; Mismatches 434; Indels 6; Gaps 1;

Qy 652 agcaactagacaaactgaatggatttcagttagtagaagaattcaccttgaagtagcctatat 711

Db 37 agccatcatgaagctgaatggccaccagtggagaaacctgcctgaaggtctctcatat 96

Qy 712 cctctgatgaatggccgccagcaaaaaccccttgccagcagcccgaggtgcgggggct 771

Db 97 cccgatgagcagatagacagagggacctgagaatggcgccgaggggcttggctctcg 156

Qy 772 tggcgagggggctctcaaggcagggggtctccaggtccggtatccagcagaacccatg 831

Db 157 gggtcagcccccgcagggctcacctgtggcagcggggcccccagcgaagcagaagt 216

Qy 832 tgattgctctgcgctgcgtggttccaccaccaatttgttggagcccatcagagaaga 891

Db 217 ggacatcccccttcggctcctggtgcccacccagtatgtgggtgccatttggcaagga 276

Qy 892 aggtgccaccattcggaaacatcaccaaacagaccagctctaaatcgatgccaccgtaa 951

Db 277 gggggccaccatccgaaacatcacaaacagaccagcccaagatagcgtgcataagaa 336

Qy 952 agaaaatgcgggggctgcgtgagaagtcgattactatctctctctctctctctctctct 1011

Db 337 ggagaaacgaggtgcagctgaaaaagccatcagtgctcactccacccttgagggtgctc 396

Qy 1012 tgcggcttgaatctattctggagattatgcataagaagcagctcaagataataaatcac 1071

Db 397 ctccgcttgaatgatcttggagattatgcataagaagcagctcaagataataaatcac 456

Qy 1072 agaagagatcccccttgaagatttttagctcataataaactttgttgagcgtcttatgttaa 1131

Db 457 tgacgaggttccccctgaagatcctggcccaataataactttgtaggcgctctcatggcaa 516

Qy 1132 agaaggaagaaattcttaaaaaaattggcgaagacacagacactaaaaatcacgatctcc 1191

Db 517 ggaaggaggaacctgaaagaggttagagcaagaaccgagacaaaaaatccactctctc 576

Qy 1192 attgcaggaattgacgctgtataatccagaacgacactattacagtttaaggaatgtga 1251

Db 577 gttgcgaagaccttaccctttacacccctgagagaccatcactgtgaagggggccatcga 636

Qy 1252 gacatgtccaaagctgagggagagatcatgaagaaaaatcaggaggtcttatgaaaaatga 1311

Db 637 gaattgtgcagggcgccgagcaggaataataatgaagaaagttcgggagccctatgagaatga 696

Qy 1312 tatgtctctatgaattcttcaagcacatttaattctcttgatttaaatctgaacgcttggg 1371

Db 697 tgtggctggccatga-----gctctcacctgacacctggcctgaacctggctgctgagg 750

Qy 1372 tctgttcccaaccacttcaggatgccacctccacatcagggccctctcagccatgac 1431

Db 751 tcttttcccaagcttcatccagcgccagtcctcccgccgctccccagcagcgcttactggggctgc 810

Qy 1432 tctctctaccccgagtttgagcaatcagaaacgagagactgttcatcagtttatccacgc 1491

Db 811 tccctatagctctcttatgcaggtcccgagcagggagagatggtgcaggtgtttatcccccgc 870

| | | | |
|----|------|---|------|
| Qy | 1492 | tctatcagtcggtgccaatcatcgcaagcagggccagacatcaagcagatttctcgctt | 1551 |
| Db | 871 | ccaggcggtggcgccaatcgcaagaggcgacatcaacagctctcccggtt | 930 |
| Qy | 1552 | tgcctggagcttcaattaaagatgctccagcggaagcaccagatgctaaagtgaagatcggt | 1611 |
| Db | 931 | tgcgcgcgctccatcaagattgcaccaccggaacacctgactccaaagtctcgatcggt | 990 |
| Qy | 1612 | gattatcactggaccaccagaggtcagttcaaggtcagggaagaatttatggaanaat | 1671 |
| Db | 991 | tatcatcactggaccgcagagcccaattcaaggtcagggaagaattctatggcaaat | 1050 |
| Qy | 1672 | taaagaagaaaacttttgactcttaagaagagaggtaaaacttgaagctcatatcagagt | 1731 |
| Db | 1051 | caaaggggagaaactctttggccccaaagggaagctgaagctggagccacatacgtcgt | 1110 |
| Qy | 1732 | gccatcctttgctgcggcagagatttatggaaaaggagcgaaaacggtgaatgaacttca | 1791 |
| Db | 1111 | gccagatcagcagctggcgggtcatigggcaaaagttgaaaaacggtgaacgagttgca | 1170 |
| Qy | 1792 | gaatttgtcaagtgcagaagtgtgtgtccctctgtaccacagacactgatgagaatgacca | 1851 |
| Db | 1171 | gaatttgcggcagctgaggtgtgtagccaagagaccagaccctgatgagaacgacca | 1230 |
| Qy | 1852 | agtgggtgtcaaaaataactggtcaacttctatgcttgcaggtgtgccagagaaaaattca | 1911 |
| Db | 1231 | ggtcatgtgaaaaatactgcgacatttctatgcccagtcagatggtctcaacgggaagatccg | 1290 |
| Qy | 1912 | ggaaattctgactcaggttaaacgagcaccaanaacag | 1948 |
| Db | 1291 | agacatccttggcccaagtttaacgacgacatcagaag | 1327 |

| | | |
|----------|--------------------|--------------------------------------|
| RESULT | 9 | |
| AAZ24600 | | |
| ID | AAZ24600 | standard; cDNA; 583 BP. |
| XX | XX | |
| AC | AAZ24600; | |
| XX | | |
| DT | 07-DEC-1999 | (first entry) |
| XX | | |
| DE | Human lung tumor | associated polynucleotide. |
| XX | | |
| KW | Human; lung tumor; | lung cancer; T cell stimulation; ss. |

AA PT PT PT PT
New isolated lung tumor polynucleotides, used to develop products for the treatment, prevention and monitoring the progression of lung cancer

PS Claim 12; Page 99; 148pp; English; English

The invention provides isolated human lung tumor nucleic acids and polypeptides. The polypeptides can be used for the treatment of lung cancer. The polypeptides and polynucleotides can be used to stimulate T

CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. the polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.
XX
XX Sequence 583 BP: 171 A; 86 C; 81 G; 235 T; 10 other;
S0

[illegible]


```

Db 158 gctcaagataataattcac--aagagatcccccttgagatttagtcataataaacttt 215
Qy 1112 gttgacgtcttattgtaagaaggaagaatacttataaaaaattgagcaagacacagac 1171
Db 216 gttgacgtcttattgtaagaaggaagaatacttataaaaaattgagcaagacacagac 275
Qy 1172 actaaatacagatatctccattgcaggaattgacgtgtataatccagaacgcactatt 1231
Db 276 actaaatacagatatctccattgcaggaattgacgtgtataatccagaacgcactatt 335
Qy 1232 acagttaaagcaatgttgagacatg-tgccaaagctgaggagagatcatgaagaaaaat 1290
Db 336 acagttaaagcaatgttgagacatgttgccaaagctgaggagagatcatgaagaaaaat 395
Qy 1291 caggagctctatgaaatgatattgtctctatgaattctcaagcacat 1339
Db 396 caggagctctatgaaatgatattgtctctatgaattctcaagcacat 444

RESULT 12
AAI56378
ID AAI56378 standard; DNA; 282 BP.
XX
AC AAI56378;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #25064 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 25064; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 282 BP; 77 A; 79 C; 70 G; 56 T; 0 other;
```

Query Match 6.7%; Score 282; DB 22; Length 282;
Best Local Similarity 100.0%; Pred. No. 7.1e-54;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 652 agcactagacaaaactgaatggtatttcagtttagagaaatttcaccttgaaaagtagcctatat 711
Db 1 agcactagacaaaactgaatggtatttcagtttagagaaatttcaccttgaaaagtagcctatat 60
Qy 712 ccctgatgaaatggcgcgcccaagcaaaaaccccttgacagagcccgaggtcgcgggggct 771
Db 61 ccctgatgaaatggcgcgcccaagcaaaaaccccttgacagagcccgaggtcgcgggggct 120
Qy 772 tgggcagaggggctctcaagcaggggtctccagagatccgtatccaagcagaaaaaccatg 831
Db 121 tgggcagaggggctctcaagcaggggtctccagagatccgtatccaagcagaaaaaccatg 180
Qy 832 tgattgctctgcctgcctgctgttccaccaccaatttggtagccatcataggaagaaaga 891
Db 181 tgattgctctgcctgctgttccaccaccaatttggtagccatcataggaagaaaga 240
Qy 892 aggtgccaccattcggaacatcaccaaacagagaccagctctaa 933
Db 241 aggtgccaccattcggaacatcaccaaacagagaccagctctaa 282

RESULT 13
AAI43398
ID AAI43398 standard; DNA; 588 BP.
XX
AC AAI43398;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #12084 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 12084; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 588 BP; 156 A; 139 C; 143 G; 150 T; 0 other;
```

Query Match 6.7%; Score 282; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 9.5e-54;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|-----------|--|--|-----|
| Qy | 652 | agcactagacaaactgaatgatttcagttcagtttagagaatttcacctgaaagtagcctatat | 711 |
| | | | |
| Db | 21 | agcactagacaaactgaatgatttcagtttagagaatttcacctgaaagtagcctatat | 80 |
| | | | |
| Qy | 712 | ccctgatgaaatgccgccagcaaaaaccccttcagcagcagcccgaggctcgccgggggct | 771 |
| | | | |
| Db | 81 | ccctgatgaaatgccgccagcaaaaaccccttcagcagcagcccgaggctcgccgggggct | 140 |
| | | | |
| Qy | 772 | tgggcagaggggctctccaagcgagggtctccaggatccgtatcccaagcagaaacccatg | 831 |
| | | | |
| Db | 141 | tgggcagaggggctctccaagcgagggtctccaggatccgtatcccaagcagaaacccatg | 200 |
| | | | |
| Qy | 832 | tgatttgctctgcgctgctgtgtccacccaatttcttgagccatcataggaagaaga | 891 |
| | | | |
| Db | 201 | tgatttgctctgcgctgctgtgtccacccaatttcttgagccatcataggaagaaga | 260 |
| | | | |
| Qy | 892 | agggtgcaccattcgggaacatcaccaaacagaccagtcctaa | 933 |
| | | | |
| Db | 261 | agggtgcaccattcgggaacatcaccaaacagaccagtcctaa | 302 |
| | | | |
| RESULT 14 | | | |
| AA02565 | | | |
| ID | AAA02565 standard; cDNA; 710 BP. | | |
| XX | XX | | |
| XX | AAA02565; | | |
| XX | XX | | |
| DT | 19-MAY-2000 (first entry) | | |
| XX | XX | | |
| DE | Human colon cancer cell line polynucleotide sequence SEQ ID NO:2556. | | |
| XX | XX | | |
| KW | Human; colon cancer; tumour; diagnosis; gene expression product; | | |
| KW | probe; detection; cancerous state; metastasis; identification; | | |
| KW | breast cancer; oestrogen receptor-positive breast cancer; therapy; | | |
| KW | oestrogen receptor-negative breast cancer; lung cancer; ss. | | |
| XX | XX | | |
| OS | Homo sapiens. | | |
| XX | XX | | |
| PN | W09958675-A2. | | |
| XX | XX | | |
| PD | 18-NOV-1999. | | |
| XX | XX | | |
| PF | 13-MAY-1999; 99WO-US10602. | | |
| XX | XX | | |
| PR | 14-MAY-1998; 98US-0085426. | | |
| PR | 15-MAY-1998; 98US-0085537. | | |
| PR | 21-MAY-1998; 98US-0085696. | | |
| PR | 21-OCT-1998; 98US-0105234. | | |
| PR | 27-OCT-1998; 98US-0105877. | | |
| XX | XX | | |
| PA | (CHIR) CHIRON CORP. | | |
| PA | (HYSE-) HYSEQ INC. | | |
| XX | XX | | |
| PI | Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J; | | |
| PI | Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A; | | |
| PI | Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I; | | |
| PI | Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B; | | |
| XX | XX | | |
| DR | WPI; 2000-126369/11. | | |
| DR | DR | | |

cancerous state of the cell from which the test sample was derived.
The polynucleotides sequences can be used in a method for detecting differentially expressed genes correlated with a cancerous state of a mammalian cell. The polynucleotides can also be used as probes for detecting and mapping related genes. They can be used in diagnosis and prognosis of diseases and disorders (e.g. identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). This is particularly for breast cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-negative breast cancer, lung cancer, and colon cancer.

Query Match 6.3%; Score 263.2; DB 21; Length 710;
Best Local Similarity 67.5%; Pred. No. 17e-49;
Matches 370; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

SQ Sequence 710 BP; 199 A; 174 C; 180 G; 152 T; 5 other;

QY 801 ctcaggatcgcgtatccaagcagaacccatgattggctctcgctcgctggtccc 860
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 105 cccctggggcacctctcaggccagacagattgattccccgtcggtactcggtccc 164

QY 861 ccgaatttgttgaggcatcataagaaagaagtgcgccaccattcggaaacataccacaac 920
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 cccagtttgttggctgatcgcgaaggaggcttgacctaaagaacatacctaagc 224

QY 921 agaccagctcaaatcgatgccaccgttaagaaaatatcggggctgcgagaagtcca 980
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 225 agaccagtcocgggttagatattccatagaaagagaactctggagctgcagagaagcgtg 284

QY 981 ttactatctctactcctgaagggcacctctgcggtttttaagtctattctggagatta 1040
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 285 tcaccatccatgccaccacagagggaacttctgaagcatgccgatgattcttgaataca 344

QY 1041 tgcataaggaagctcagagataaaattccagaagaagatcccccttgaaatttagctc 1100
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 345 tgcgaagagcagatgagaccaaacctagccgaagagattcctctgaaatcttggcac 404

QY 1101 ataataactttgtggacgtcttatgttaaagaggaagaatacttaaaaaattgagc 1160
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 405 acaatggttggtagacgtgattggaaagaagcagaaaatttgagaaaattgaaac 464

QY 1161 aagcacagacactaaaatcacgatatctccattgcagggaattgcacgtgtataatccag 1220
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 465 atgaaacaggaccagaataacaatctcatcttctgcaggatttgagcatatacaaccgg 524

QY 1221 aagcactattacagtttaaaggcaatgttagacatgtgccaaagctcgaggaggagatca 1280
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 525 aaagaacatcactgtgaaggcagctgaggcctgtgccagctgctgagatagagatta 584

QY 1281 tgaagaaaatcaggagctttatgaaaatgatattgcttatgaacttccaagcacatt 1340
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 585 tgaagaactcgtgaggcccttgaaaatgatatgctggctgttaacgtaagtcctcaa 644

QY 1341 taattcct 1348
 ||| |||
DB 645 tgcttct 652

RESULT 15
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX AC AAF58252;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2002, 14:54:23 ; Search time 19840.1 Seconds
(without alignments)
2264.506 Million cell updates/sec

Title: US-09-685-696-175

Perfect score: 4181

Sequence: 1 ggtgatgcgttgggtgtg.....aaaaaaaaaaaaaaaaaaaa 4181

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 1680.6 | 40.2 | 2202 | AK011689 | AK011689 Mus muscu |
| 2 | 842 | 20.1 | 904 | AU117945 | AU117945 AU117945 |
| 3 | 802 | 19.2 | 876 | AU142785 | AU142785 AU142785 |
| 4 | 801.2 | 19.2 | 2780 | AK013940 | AK013940 Mus muscu |
| 5 | 759.2 | 18.2 | 808 | AU117427 | AU117427 AU117427 |
| 6 | 723.6 | 17.3 | 745 | AU134323 | AU134323 AU134323 |
| 7 | 716.2 | 17.1 | 902 | BE869656 | BE869656 601446879 |
| 8 | 648.4 | 15.5 | 937 | BI090231 | BI090231 602857184 |
| 9 | 638.8 | 15.3 | 733 | AU127673 | AU127673 AU127673 |
| 10 | 634 | 15.2 | 751 | AU139887 | AU139887 AU139887 |
| 11 | 630.4 | 15.1 | 790 | BE545535 | BE545535 601070391 |
| 12 | 629.6 | 15.1 | 816 | AU139815 | AU139815 AU139815 |

| | | | | | |
|----|-------|------|------|----|----------|
| 13 | 617.6 | 14.8 | 805 | 11 | BG108548 |
| 14 | 617 | 14.8 | 769 | 11 | BF608004 |
| 15 | 610.6 | 14.6 | 633 | 11 | BG499373 |
| 16 | 606.4 | 14.5 | 1101 | 11 | BG576431 |
| 17 | 601.2 | 14.4 | 1005 | 11 | BG435192 |
| 18 | 588.2 | 14.1 | 940 | 11 | BF034031 |
| 19 | 578.8 | 13.8 | 754 | 11 | BF215050 |
| 20 | 573.8 | 13.7 | 643 | 10 | AU134724 |
| 21 | 572.8 | 13.7 | 590 | 11 | BG434352 |
| 22 | 570.5 | 13.6 | 662 | 11 | BF185290 |
| 23 | 570.2 | 13.6 | 654 | 11 | BF185289 |
| 24 | 567 | 13.6 | 715 | 11 | BG069090 |
| 25 | 545.8 | 13.1 | 565 | 10 | AI807368 |
| 26 | 536.8 | 12.8 | 975 | 11 | BG115140 |
| 27 | 531.8 | 12.7 | 580 | 10 | AI086619 |
| 28 | 530.4 | 12.7 | 640 | 10 | AW958019 |
| 29 | 529 | 12.7 | 898 | 11 | BF664818 |
| 30 | 528.6 | 12.6 | 575 | 10 | AW044616 |
| 31 | 524.6 | 12.5 | 766 | 11 | BG434285 |
| 32 | 518.8 | 12.4 | 567 | 10 | AW628546 |
| 33 | 514 | 12.3 | 587 | 11 | BG079106 |
| 34 | 509.6 | 12.2 | 639 | 11 | W81035 |
| 35 | 506 | 12.1 | 1144 | 11 | BG531972 |
| 36 | 505.6 | 12.1 | 554 | 11 | D63197 |
| 37 | 500.6 | 12.0 | 547 | 10 | AW043682 |
| 38 | 495 | 11.8 | 495 | 10 | AI088047 |
| 39 | 495 | 11.8 | 645 | 10 | AA163596 |
| 40 | 494.6 | 11.8 | 501 | 10 | AI824109 |
| 41 | 493.2 | 11.8 | 573 | 11 | BG082113 |
| 42 | 492.4 | 11.8 | 847 | 11 | BF697640 |
| 43 | 489 | 11.7 | 490 | 10 | AI424833 |
| 44 | 488 | 11.7 | 712 | 11 | BG528016 |
| 45 | 487.2 | 11.7 | 907 | 11 | BF697335 |

ALIGNMENTS

RESULT 1

| LOCUS | AK011689 | 2202 bp | mRNA | HTC | 05-JUL-2001 |
|------------|---|---------|------|-----|-------------|
| DEFINITION | Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610036B18, full insert sequence. | | | | |
| ACCESSION | AK011689 | | | | |
| VERSION | AK011689.1 GI:12847972 | | | | |
| KEYWORDS | CAP trapper. | | | | |
| SOURCE | Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone:2610036B18. | | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | | | |
| AUTHORS | 1 (bases 1 to 2202) | | | | |
| TITLE | Carninci,P. and Hayashizaki,Y. | | | | |
| JOURNAL | High-efficiency full-length cDNA cloning | | | | |
| MEDLINE | Methods in enzymology. 303, 19-44 (1999) | | | | |
| PUBMED | 99279253 | | | | |
| REFERENCE | 10349636 | | | | |
| AUTHORS | 2 (bases 1 to 2202) | | | | |
| TITLE | Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. | | | | |
| JOURNAL | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes | | | | |
| MEDLINE | Genome research. 10 (10), 1617-1630 (2000) | | | | |
| PUBMED | 20499374 | | | | |
| REFERENCE | 11042159 | | | | |
| AUTHORS | 3 (bases 1 to 2202) | | | | |
| TITLE | Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakauchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue.K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., | | | | |

| | |
|----------|------------|
| BG108548 | 602278519 |
| BF608004 | MX1_00094 |
| BG499373 | 602547471 |
| BG576431 | 602595673 |
| BG435192 | 602507712 |
| BF034031 | 601456355 |
| BF215050 | 601847395 |
| AU134724 | AU134724 |
| BG434352 | 602506351 |
| BF185290 | 601843593 |
| BF185289 | 601843592 |
| BG069090 | H3072F04 - |
| AI807368 | W47803.X |
| BG115140 | 602316003 |
| AI086619 | Oz76f02.X |
| AW958019 | EST370089 |
| BF664818 | 602117746 |
| AW044616 | WY78B05.X |
| BG434285 | 602506251 |
| AW628546 | h141a12.X |
| BG079106 | H3035G11 - |
| W81035 | zd84c01.r1 |
| BG531972 | 602560575 |
| D63197 | HUM504G10B |
| AW043682 | WY80a10.X |
| AI088047 | OW60h07.X |
| AA163596 | mn38g02.r |
| AI824109 | w146G11.X |
| BG082113 | H3072F04 - |
| BF697640 | 602131173 |
| AI424833 | tg37e07.X |
| BG528016 | 602565989 |
| BF697335 | 602129877 |

Db 181 AGGCAAAACGGTGAATGAACTTCAGAAATTGTCAAGTGCAGAGATTGTTGTCCCTCGTGA 240


```
QY 547 ttactagtcacagatgagagtggtgagagctgtgagcaagtgaacacactgcgaaac 606
   || || || || || || || || || || || || || || || || || || || || ||
Db 575 CTTGCTGGCTCAGTACGTGAGAACTGTGAGCAAGTCAACACTGAAAGTGAGAC 634

QY 607 tgcagttgtaaatgtaacattccagtaagagcgaagctagacaagcactagacaaact 666
   || || || || || || || || || || || || || || || || || || || || ||
Db 635 AGTGTGGTCAACGTCACCTTACTTAACCGGAGCAGACAGCAAGCTATCATGAAGCT 694

QY 667 gaatgatttcagttagagaatttcacctgaaagttagcttatccctgataaagtgc 726
   || || || || || || || || || || || || || || || || || || || || ||
Db 695 AATGGCCATCAACTGAGAACCATGCCCTGAAGGCTCTCTACATACCTGATGAGCAGAT 754

QY 727 cgcacgaacaaaccccttgagcagccccgaggtcgccgggggcttggcgaggggctc 786
   || || || || || || || || || || || || || || || || || || || || ||
Db 755 AACGCAAGTCTCTGAGAACTGGGCTCTGTGAGGCTTTTGGGTCTCGGGGCCAGCCCGCA 814

QY 787 ctcaagcaggggtctccaggtaccgtatccaaagcagaacacatgtgattgctctgcg 846
   || || || || || || || || || || || || || || || || || || || || ||
Db 815 AGGTCGCCGCTGGCAGCAGGGGCTCCAGCCAAAGCAGCAGCCAGTGGACATCCCTCTCCG 874

QY 847 cctgctgggtcccccaccacaaattgttgagccatcatagaaagaagggtgcccaccattcg 906
   || || || || || || || || || || || || || || || || || || || || ||
Db 875 GTCCTGGTGCCTACGCAGTATGTAGGCGCTATCATTTGCCAAGGAGGTGTCACCATCCG 934

QY 907 gaacatcaccaaacagaccacagtcataaaatcgatgtccaccgtataaagaatacgggggc 966
   || || || || || || || || || || || || || || || || || || || || ||
Db 935 AAACATCACAAAACAGACGCGAGTCCAAAATAGACGTGCATAGGAAGAGAAATCGCGGCGC 994

QY 967 tgcctgagaagtcgattactatctctactcctgaagcaccctctgcggcttgtaagtc 1026
   || || || || || || || || || || || || || || || || || || || || ||
Db 995 TCGGGAAGAGGCCATCAGCGTGTGATTCACCCCTGAAGCGTCTCTCCGCGTGCAGAAAT 1054

QY 1027 tatctgagagattatcataagaagcctcaagatatataaattcacagaagagatccctt 1086
   || || || || || || || || || || || || || || || || || || || || ||
Db 1055 GATCTTGGAGATTATGCANAGAGAGCAAGGACACACAAACGGCAGATGAAGTTCCCT 1114

QY 1087 gaagatttagctcataataactttgttgagctcttattggttaaagaaggaaatct 1146
   || || || || || || || || || || || || || || || || || || || || ||
Db 1115 GAAGATCTCTGGCTCATAAACATCTCTCGGCGACTCATTTGGCAAGAGGGCGGAACCT 1174

QY 1147 taaaaaattgacagaacacagacactaaaatcacagatatctcccttgcaggaattgac 1206
   || || || || || || || || || || || || || || || || || || || || ||
Db 1175 GAAGAAGTGGAGCAGACAGACAGAGAGATGATCAGCATCTCATCGCTCCAGGACCTCAG 1234

QY 1207 gctgtataatccagaacgcaactattacagttaaaggcaattgttgagacatgtgccaagc 1266
   || || || || || || || || || || || || || || || || || || || || ||
Db 1235 GCTCTATAACCTTGAGAGNCCATCACTGTGAAGGGCGGCTTACGAGAACGCTGGCGCCATGAG 1294

QY 1267 tgaggaggagatcatgaagaaatcaggagtgcttattgaaaatgatattgcttctatgaa 1326
   || || || || || || || || || || || || || || || || || || || || ||
Db 1295 CGAGCAGGAGATCATGAAGAACTTCGAGAGGCTTACGAGAACGCTGGCGCCCATGAG 1354

QY 1327 tctcaagcacatttaattccttgattaaatctgaacgcttgggtctgttccccaccac 1386
   || || || || || || || || || || || || || || || || || || || || ||
Db 1355 CTGTGACCTCCCACTCATCTCTGGGCTTAACCTTGGCTGTGTAGGCTCTTCCCAAGCTTC 1414

QY 1387 ttcagagatgccaactcccacctcagggcccccttcagcatgaactccctcaccgcga 1446
   || || || || || || || || || || || || || || || || || || || || ||
Db 1415 ATCCAGCGTGTCTCTCTCCAGCAGAGTGTACACGGGGCTGTCTCCCTATATAGTCTCTT 1474

QY 1447 gtttgagcaatcagaacacgagctgttcatcagtttattccccagctctcatcagtcggtgc 1506
   || || || || || || || || || || || || || || || || || || || || ||
Db 1475 CATGCAGGCTCCGAGCAGAGATGCTACAAGTGTTCATCCCGCCCGCAGGCTGTGGGCGC 1534

QY 1507 catcatcgcgaagcagggccagacacatcaagcagcttctcgtcttgcgtggagcttcaat 1566
   || || || || || || || || || || || || || || || || || || || || ||
Db 1535 CATCATTTGGCAAGAGGGCCAGCACATAAACAACTCTCCCGCTTCGCCAGCGCTCCCAT 1594

QY 1567 taagattgctccagcggagcagcagatgctaaagtggagtggtgattatcaactgacc 1626
   || || || || || || || || || || || || || || || || || || || || ||
Db 1595 CRAAGATTGCACCACCAACACCTGACTCCAAAGTTCGAATGGTCTCATCACTGAGACC 1654
```

```
QY 1627 accagaggtcagttcaaggtcagggaggaatttatggaaaaataaaagaagaaactt 1686
   || || || || || || || || || || || || || || || || || || || || ||
Db 1655 CCAGAGGCTCAGTTCAAGCCCGCAGGGAAGAAATCTATGGCAAACTAAAGAGAATTT 1714

QY 1687 tgttagtctctaaagaagaggtgaaacttgaagctcatatcagagtgccatcccttctgctg 1746
   || || || || || || || || || || || || || || || || || || || || ||
Db 1715 CTTTGGTCCCAAGGAGGAAGTAAGCTAGAGACCCACATACGGGTTCCTCCGCTTCAGCAGC 1774

QY 1747 tggcagagttatttgaaaggagggcaaacggtgaaatgaacttcagaatttgcacagtgc 1806
   || || || || || || || || || || || || || || || || || || || || ||
Db 1775 CGCCCGTGTCTCGGCAAGAGCGGCAAAACGGTGAATGACCTGCAGAACTTGCACCGCAGC 1834

QY 1807 agaagttgttgcctcctgtgaccagacacctgatgagaatgaccagaagtgtgtcctcaaat 1866
   || || || || || || || || || || || || || || || || || || || || ||
Db 1835 TGAGGTGGTAGTGCCAAAGAGACAGACACCCCGGATGAGAACGACCAAGTCATGTTAAGAT 1894

QY 1867 aactggtcactcttatgcttgcaggttgcgccagagaaaaattcaggaaaattctgactca 1926
   || || || || || || || || || || || || || || || || || || || || ||
Db 1895 CATCGGACATTTTATGCCAGCCAGATGGCTCAGCGGAAGATCCGAGACATCTCTGGCTCA 1954

QY 1927 ggtaaaagcagcaccacaacag 1948
   || || || || || || || || || || || || || || || || || || || || ||
Db 1955 AGTTAAGCACACACACAGAGAG 1976

RESULT 5
AUI17427 LOCUS AUI17427 808 bp mRNA EST 19-OCT-2000
DEFINITION AUI17427 HEMBA1 Homo sapiens cDNA clone HEMBA1001372 5', mRNA
sequence.
ACCESSION AUI17427
VERSION AUI17427.1 GI:10932395
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 808)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute: cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1. .808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1001372"
/tissue_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18SFL3"
BASE COUNT 262 a 127 c 160 g 255 t 4 others
ORIGIN
|||||
Query Match 18.2%; Score 759.2; DB 10; Length 808;
Best Local Similarity 97.1%; Pred. No. 2.9e-111;
Matches 782; Conservative 0; Mismatches 21; Indels 2; Gaps 1;
QY 2367 tagattgctctctaaatcccaattgttaaaatttgatcagataataattatcacagaacttaa 2426
|||||
```

| | | | |
|------------|---|---|------------|
| Db | 1 | TAGATTGCTCTTAATCAATTTGTTAAATTTGGATCAGAATAATATACAGGAACCTTAA | 60 |
| Qy | 2427 | atgttaagccattagcatagaaaaactgtctcagtttttatttttacctaacaactaacat | 2486 |
| Db | 61 | ATGTTAAGCCATTAGCATAGAAAAACTGTTCTCAGTTTATTTTACCTAACACTAACAT | 120 |
| Qy | 2487 | gagtaacctaaaggaagtcgaatggtgtgttgccaggggtattaaacgtgatttttact | 2546 |
| Db | 121 | GAGTAACCTAAGGGAAGTGCATAGTGGTGGCAGGGGTATTAAACGTGATTTTACT | 180 |
| Qy | 2547 | caactacctcaggtattcagtaatacaatgaaaagcaaaatgtctctttttttgaaaa | 2606 |
| Db | 181 | CAACTACCTCAGGTATTTCAGTAAATACATGAAGAAGCAAAATGTTCTTTTTTGA | 240 |
| Qy | 2607 | ttttatacttttataatgatagaagtcacacccgtttttttaaataaaatattt | 2666 |
| Db | 241 | TTTTATATACTTTATAATGATAGAAGTCCACCGTTTTTTTAAAAATAAATTTAA | 300 |
| Qy | 2667 | aacagcaatcagctaaacaggcaaatgaatttttacttctggtggtgacagtaaaagt | 2726 |
| Db | 301 | AACAGCAATCAGCTAACAGGCAAAATTAAGATTTTACTTCTGGCTGTGACAGTAA | 360 |
| Qy | 2727 | ggaaaattcaatttcaggggttttttgaggtcttttgacacagttattagttaaatca | 2786 |
| Db | 361 | GGAAAATTAATTTTTCAGGGTTTTTTGAGGCTTTTGACACAGTTATTAGTTAA | 420 |
| Qy | 2787 | tcaaaaaacagagcagtcgcttagtatctggagagcagcactaccatttattcttc | 2846 |
| Db | 421 | TCAAAAATACGAGCAGTGCCTAGTATCTGAGAGCAGCAGTACCATTATTTCT | 480 |
| Qy | 2847 | tatagttgggaaagttttttgacgttactaacaagaagtgctgcagagagattttg | 2906 |
| Db | 481 | TATAGTTGGGAAAGTTTTTGTGCGGTACTTAACAAAGTGTGTCAGGAGATTTT | 540 |
| Qy | 2907 | tggtttaaatgcttcaagagacacttcagttttttgttttagctacatgattgaatg | 2966 |
| Db | 541 | TGGTTTAAATGGCTTCAGGAGACTTCAGTTTTTTGNTTAGCTACATGATTGA | 600 |
| Qy | 2967 | taaatgtttgtctctgactatcaatcacctaaagaagtgcatcagtgagagatgca | 3026 |
| Db | 601 | TAAATGCTTTTGCTTCGACTATCAATCAATACCTAAGAAGTGATCAGTGAAG | 660 |
| Qy | 3027 | agactttcaactgactggtgcaaaagcaagcttttagcttctttaggatgcttagttg | 3086 |
| Db | 661 | AGACTTTCAACTGACTGTGCAAAAGCAAGCTTTAGCTTTGGCTTATAGGATGCT | 720 |
| Qy | 3087 | ccactacattcagaccatgggacagtcataagtggtgtgacagtggtttaaac-- | 3144 |
| Db | 721 | CCACTACACTTTAGACCAATGGGACAGCATANATGTGGGGGCTGGGGTTTAA | 780 |
| Qy | 3145 | caaaaggctacatttccatggggcc | 3169 |
| Db | 781 | CAAAAGGCTACATTTCCATGGGGC | 805 |
| RESULT | 6 | | |
| AUI34323 | | | |
| LOCUS | AUI34323 | 745 bp | mRNA |
| DEFINITION | AUI34323 | OVARC1 Homo sapiens | cdna clone |
| ACCESSION | AUI34323 | | |
| VERSION | AUI34323.1 | | |
| KEYWORDS | EST | | |
| SOURCE | human | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| | Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., | | |
| | Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and | | |
| | Isogai,T. | | |
| TITLE | HRI human cdna project | | |
| JOURNAL | Unpublished (2000) | | |

| | | | | |
|---------------------------|---|-------|-------|----------------|
| COMMENT | Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1..745 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="OVARC1001723" /clone_lib="OVARC1" /tissue_type="ovary, tumor tissue" /note="Vector: pME18SFLJ" | | | |
| BASE COUNT | 232 a | 168 c | 170 g | 172 t 3 others |
| ORIGIN | | | | |
| Query Match | 17.3%; Score 723.6; DB 10; Length 745; | | | |
| Best Local Similarity | 98.9%; Pred. No. 1.4e-105; | | | |
| Matches 737; Conservative | 0; Mismatches 7; Indels 1; Gaps 1; | | | |
| QY | 1455 aatcagaacgagacgtgttcacatcagtttattccagctctatccagctcagtcggtgccatcagc | 1514 | | |
| Db | 1 AATCAGAAACGGAGACTGTTTCATCTGTTTATCCAGCTCTATCAGTCGTGTCATCATCG | 60 | | |
| QY | 1515 gcaagcagggccagcacatcaagcagcttctcgtcttgctgagcttcaatgaattg | 1574 | | |
| Db | 61 GCAAGCAGGGCCAGCACATCAAGAGCTTCTCGCTTTCGTGAGCTTCAATTAAGATTG | 120 | | |
| QY | 1575 ctccagcggaaagcaccagatgctaaaagtgaggtgaggtattctactgagcaccacagagg | 1634 | | |
| Db | 121 CTCACGCGGAAGCAGATGTTAAAGTGAGGATGGTATTATCAGTCGACCCAGGAGG | 180 | | |
| QY | 1635 ctcaagttcaaggtcagggaagaatttgaaaaattaaagaagaactttgttagtc | 1694 | | |
| Db | 181 CTCAGTTCAAGGCTCAGGGAAGAAATTTATGGAAATTTAAAGAAAGAAACTTTGTAGTC | 240 | | |
| QY | 1695 ctaaaagaaggtgaaacttgaaagctcatcagagtgccatcccttgcgtgcgcagag | 1754 | | |
| Db | 241 CTAAAGAGAGGTGAAACTTTGAAGCTCATATCAGAGTGCCATCTTGTGCTGGCAGAG | 300 | | |
| QY | 1755 ttattggaaaaagggcaaaacgggtgaatgaaacttcagaatttgcgaagtgcagaagtgc | 1814 | | |
| Db | 301 TTATTTGAAAAAGGAGGCAAAACGGTGAATGAATTTGTAAGTTCAGAAATTTGTAAGT | 360 | | |
| QY | 1815 ttgtccctcgtagccagacacctgatgagaatgacaaagtgtgtcagaaaataactggtc | 1874 | | |
| Db | 361 TTGTCCCTCGTGACCAGACACCTGATGAGATGACCAAGTGGTTGTCAAAATAACTTGGTC | 420 | | |
| QY | 1875 actcttatgtctgcccaggttgcccagagaaaaattcagaaaaattctgactcaggtgaaagc | 1934 | | |
| Db | 421 ACTTCTATGCTTGCCAGGTTGCCAGAGAGAAAAATTCAGGAAATTTCTGACTCAGGTAAAGC | 480 | | |
| QY | 1935 agcaccacaacacagaagcgtctgcaaaagtggaccacctcagtcagagacggaagtgaagc | 1994 | | |
| Db | 481 AGCACCAACACAGAAAGGCTCTGCAAAGTGGACCACTCAGTCAAGACGGAAGTAAGGC | 540 | | |
| QY | 1995 tcaggaacacagccaccacagagcagatgccaaacaaagacagattgtcttaaccaaca | 2054 | | |
| Db | 541 TCAGGAACAGGCCACACAGAGGCGAGATGCCAAACCAAGACAGATTGCTTAAACCAACA | 600 | | |
| QY | 2055 gatggcgtgacccctctatccagaatcagtcacagcaagtttttactagcagcagttggtt | 2114 | | |
| Db | 601 GATGGCGCTGACCCCTATCCAGAAATCATCATGCAAACTTTTACTAGCCAGTTGTTT | 660 | | |
| QY | 2115 ctgaggaccagggcaacttttgaaactcctgtctctgtgtgagaatgtatacctttatgctctct | 2174 | | |

```

Db 661 CTGAGGACCAGGCAACTTTTGAACCTCTGNCCTCTGTGANAATGTAT-CTTTATGCTCTCT 719
QY 2175 gaaatgtatgacacccagctttaa 2199
|||||
Db 720 GAAATGTATGACNCCCACTTTAA 744

RESULT 7
BE869656
LOCUS 601446879F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850963 5',
DEFINITION mRNA sequence.
ACCESSION BE869656
VERSION BE869656.1 GI:10318536
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 902)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9570 row: n column: 20
High quality sequence stop: 683.
Location/Qualifiers
1. .902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3850963"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 297 a 187 c 210 g 208 t
ORIGIN

Query Match 17.1%; Score 716.2; DB 11; Length 902;
Best Local Similarity 92.7%; Pred. No. 1.8e-104;
Matches 776; Conservative 0; Mismatches 53; Indels 8; Gaps 2;

QY 1509 tcatcggaagcaggccagcaatacaagcagcttctcgttgcgtgagcttcaatta 1568
|||||
Db 1 TCATCGGCAAGCAGGCGCAGCACATCAAGCAGCTTTCTCGCTTGTGGAGCTTCAATTA 60

QY 1569 agattgtccagcggagcaccagatgctaaagtggatggattacactggaccac 1628
|||||
Db 61 AGATTCTCCAGCGGAGCACCAGATGCTAAAGTAGAGTGTGATATATCATCTGGACCAC 120

QY 1629 cagaggtcagttcaggggtcaggggaagtatttgaaaaataagaagaaactttg 1688
|||||
Db 121 CAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGAAAAATTAAGAAGAAACCTTG 180

QY 1689 ttagtctaaagaagagggtgaacttgagctcattatcagatgacctttgtctgtcgtg 1748
|||||
Db 181 TTAGTCTCTAAAGAAGAGGTGAACACTTGAAGCTCATATCAGAGTGCCTCTTGTGCTG 240

QY 1749 gcagagttattgaaaaagagcaaacggtgaatgaacttcagaatttgcgaagtgcag 1808
|||||
Db 241 GCAGAGTATTGGAAGAGGAGGCAAAACGCTGAATGAACCTTCAGAAATTTGTCAAGTGCAG 300

```

```

QY 1809 aagttgttctccctcgtgaccagacacctgatgagaatgaccaagtgtgttcaataa 1868
|||||
Db 301 AAGTTGTGTCCCTCGTGACCAGACACCTGATGAGAATGACCAAGTGTGTCAATAA 360
|||||
QY 1869 ctggtcacttctatgcttgcagggttcccagagaaaaattcaggaattctgactcagg 1928
|||||
Db 361 CTGGTCACTTCTATGTCTGGCAGGTTGCCAGAGAAAATTCAGGAAATTCGACTCAGG 420
|||||
QY 1929 taaagcagcaccacaacagaagcctctgaaaagtggaccacctcagtcgaagcggaggt 1988
|||||
Db 421 TAAAGCAGCACCACACAGAGGCTCTGCAAGTGGACCACCTCAGTCAAGACGGAAGT 480
|||||
QY 1989 aaaggctcaggaacacccaccacagagcagatgcacaaacacacacagattgcttaa 2048
|||||
Db 481 AAAGGCTCAGGAAACGCCACCACAGAGGATGCCAAACCAACACAGATTTGCTTAA 540
|||||
QY 2049 ccaacagatggcgctgacccccctatccagaatcacatgacaaagttttacctagcag 2108
|||||
Db 541 CCAACAGATGGCGCTGACCCCTATCCAGAATCACATGCACAAAGTTTACCTTAGCCAG 600
|||||
QY 2109 ttgtttcaggaccagcgaacttttgaactcctgtctctgtgagaaatgtatactttatg 2168
|||||
Db 601 TTGTTTCTGAGGACCAGCAACTTTTGAACCTCTCTCTGTGAAAATGTATATCTTATG 660
|||||
QY 2169 ctctcgaatgtatgacacccagctttaaacaacaaacacaaacacaaacacaaaggg 2228
|||||
Db 661 CTCTCTGAATGTATGACAAACACAGCCCTTAA---AAACCAACAAATAAGAACAAAGG 717
|||||
QY 2229 ggggaggaggaagagagagagagctgcacttccctt-----gtttagtctcacag 2283
|||||
Db 718 GGGGGAGGAGCGCAGAGAGAAAGCTTGGAAATTCCTTTGTGGTGTAAATCTCACAGG 777
|||||
QY 2284 tatacacgatattctaatcttcttaattcccccaataatgccagaattggccta 2340
|||||
Db 778 ATAACAGAAATATCTCTTAATTAATTTCCATAATAGGCCGCAAAATGGGTAA 834
|||||

```

```

RESULT 8
BI090231 937 bp mRNA EST 20-JUN-2001
LOCUS 602857184F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4998572 5',
DEFINITION mRNA sequence.
ACCESSION BI090231
VERSION BI090231.1 GI:14508561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 937)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11027 row: g column: 21
High quality sequence stop: 758.
Location/Qualifiers
1. .937
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4998572"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
FEATURES
source

```


Db 361 ATCTCCATTGAGGAAATGACGCTGTATATATCCAGACGCACTATTACAGCTTAAGGCAA 420
QY 1246 tgttgagacatgtgcaaaagctgaggagatcatgaagaataatcaggagctctatga 1305
Db 421 TGTGTAGACATGTGCCAAAGCTGAGGAGAGATCATGAAGAAATACAGGAGCTTTATGA 480
QY 1306 aaatgatattgcttctatgaattcaagcacacatttaattcctcggattaaatctgaacgc 1365
Db 481 AAATGATATTGCTTCTATGAATCTTCNAGCACATNTAATTCCTGNAFTAAATCTGAACGC 540
QY 1366 ctltgggtctgttcccccacttcaaggatgccacctccacactcagggcccttcacgc 1425
Db 541 CTGCGTCTGTTCCACCCACTTCAGGATGCCACCTCCACCTCAGGCGCCCTTCAGC 600
QY 1426 catgactcctccctaccgcagtttgagcaatcagaacaggagactgttcatcagtttat 1485
Db 601 CATGACTCCTCCCTNCCGCGAGTTTGAGCAATCCGAACGAGACTGGTTCNTCTGTTT 660
QY 1486 ---cccagctctat-cagtcgggtgcaatcagg 1515
Db 661 ATCCCCACCTCTATCCAGTCGGTGGCCATCATCCG 694

RESULT 10
AUI39887 751 bp mRNA EST 25-OCT-2000
LOCUS AUI39887 PLACE1 Homo sapiens cDNA clone PLACE1011527 5', mRNA
DEFINITION sequence.
ACCESSION AUI39887
VERSION AUI39887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
TITLE HRI human cDNA project
JOURNAL HRI human cDNA project
COMMENT Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES
Source Location/Qualifiers
1..751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1011527"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SPL3"
BASE COUNT 253 a 131 c 137 g 226 t 4 others
ORIGIN

Query Match 15.2%; Score 634; DB 10; Length 751;
Best Local Similarity 94.5%; Pred. No. 2.3e-91;
Matches 718; Conservative 0; Mismatches 30; Indels 12; Gaps 6;
QY 2977 gtgcttctgactatcaataacacaaagtgatcagtgaaagatgcaagactttcaa 3036
Db 1 GTGCTTCTGACTATCAATACCTAAAAAAGTGCAATCAGTCAAGAGATGCAAGACTTTCAA 60
QY 3037 ctgactggcaaaaagcaagcttagctgtcttataggatgcttagttgcccactacact 3096

Db 61 CTGACTGGCAAAAGCAAGCTTTAGCTTGCTTATAGATGCTTTAGTTGCCATACACT 120
QY 3097 tcagaccaatgggacagctcatagatgggtgacagtggttttaaacgcacaaaagcctaca 3156
Db 121 TCAGACCAATGGGACAGTCATAGATGGTGTGACAGTGTTTAAAGCAACAAAAGGCTACA 180
QY 3157 ttccatggggccagcactgcatgagcctcactaaagctattttgaagatttttaaacac 3216
Db 181 TTTCCATGGGGCCAGCACTGTCATGAGCCTCCTCAAGCTATTTTGAAGATTTTAAAGCAC 240
QY 3217 tgataaataaaaaaaataaaataatagactccaccttaagtagtaagataaacag 3276
Db 241 TGATAAAT-----AAAAAATAAATAAGACTCCACCTTAAGTAGTAAGTATAACAG 294
QY 3277 gattctgtatactgtgcaatcagttcttgaaaaaaagctcaaaagatagagaatacaa 3336
Db 295 GATTTCTGTATACTGTGCAATCAGTCTTTTGAAAAAATAAGTCAAAAGATAGAGAAATACA 354
QY 3337 gaaaagtttngggatataatttgaatgactgtgaaacatatgacctttgataaacgac 3396
Db 355 GAAAAGTTTTTGGGATATAAATTTGAATGACTGTCAAAACATATGACCTTTGATAACGAAC 414
QY 3397 tcatttgcactccttgacagcaaaagccagtcagcaaatgtgtgtgggtgtgggtgg 3456
Db 415 TCATTTGCTCCTCTGTGACAGCAAAAGCCAGTACAGTGTGTGGGTGTGGGTGG 474
QY 3457 tctccaagggccagcgtctctgaaattgttttggagtttggnttgnaaagatgata 3516
Db 475 TCTCCAAAGGCCACGCTCTCTCGAATGATTTTGTGAGTGTGTTTCCGAAAGAAC 533
QY 3517 cagncatgttcactgactgtttnaaggacatatntataaccctttaaataaaatcccc 3576
Db 534 CAGTCATGTTTACACTGATC-TAAAGGACATATA-TATAACCTTTTAAAAAAGAAATCAC 591
QY 3577 tgctcattcttcttcogagatgaatttcogataca-gactagatgtcttcctgaagata 3635
Db 592 TGCTCATCTCTTATTTCAAGATGAATTTCTATACAGACTAGATGTGTTTCCGAAAGAAC 651
QY 3636 attagacattntgaaaatgatttaaaagtggttttccttaagtctctgaaaaaagtttc 3695
Db 652 ATTAGACATTTTGAANAATGATTTAAAGTGTTTTCCCTTAANGTCTCCGAAACAAAGTTTC 711
QY 3696 -tttctgagttttaaccacaaaagtcgctcttttgcac 3733
Db 712 CNTTTGGAGGTTTAAACCCAAAAAAGTGCCCTTTTGNCAAC 751

RESULT 11
BE545535 790 bp mRNA EST 09-AUG-2000
LOCUS 601070391F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456407 5', mRNA sequence.
DEFINITION BE545535
ACCESSION BE545535
VERSION BE545535.1 GI:9774180
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 790)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov


```
Db 470 AAAAAATTAATTTAAATTTAACAGCAATCAGCTAACAGGCAAAATTAAGATTTTTACTTCT 529
QY 2708 ggctggtagcagtaaaagctggaataatttcagggttttttgaggcttttgacacagt 2767
Db 530 GGCTGGTGACCGTAAGCTGGAATAATTAATTCAGGGTTTTTTTGAGGCTTTTGACACAGT 589
QY 2768 tattagttaaatcaaatgtttcaaaaatacaggagcagtcgctagtatatcgtgagcgagcac 2827
Db 590 TATTAGTTAAATCAAAATGTTTCAAAAATACGAGCAGTCGCTAATATCTGGAGAGCAGCAC 649
QY 2828 taccattattcttccattcat-agtgggaagattttgacggtactaacaagaagtggt- 2885
Db 650 TACCAATTATCTTCTTATTAAGTTGGGAAAGTTTTCACCGGTACTACAAAGTGGTC 709
QY 2886 cgcagagattttggaacgctggtttaaatggtctc-agagacttcagttttttgttt 2944
Db 710 CGCAGGAGATTTTGGACGCTGGTGTAAATGCTTCAAGGAGACTTCAGTTTTTTGGTT 769
QY 2945 agc--tacatgattgaatgcataataaaagtctttgtg 2979
Db 770 ANNCTACAATGATTGAATGCATATAAATGCNTTGGG 806

RESULT 13
BG108548 805 bp mRNA EST 30-JAN-2001
LOCUS 602278519f1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4366085 5',
DEFINITION mRNA sequence.
ACCESSION BG108548
VERSION BG108548.1 GI:12602394
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10016 row: f column: 06
High quality sequence stop: 722.
FEATURES
Location/Qualifiers
1..805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4366085"
/clone_lib="NIH_MGC_86"
/tissue_type="Osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: bone; vector: pCMV-SPORT6; site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 242 a 130 c 136 g 297 t
ORIGIN
```

Query Match 14.8%; Score 617.6; DB 11; Length 805;

Best Local Similarity 90.9%; Pred. No. 8.9e-89;

Matches 728; Conservative 0; Mismatches 62; Indels 11; Gaps 7;

```
QY 3280 ttctgtactgtgcatacgtttttgaaagaaagtcaaaagatagagaata-caaga 3338
|||||
```

```
Db 5 TTCTGTATACTGTGCAATCAGTTCTTTTGAAAAAAGCTAAAAAGATAGAGAATACCAAGA 64
QY 3339 aaagttttnggataataatttgaaactgactgaaacatatgacccctttgataacgaactc 3398
|||||
Db 65 AAAGTTTTTGGGATATAAATTTGAATGACTGTGAAAAACATATGACCTTTTGATAACGAATC 124
QY 3399 atttgcctaccccttgacagcaaaagcccgactacgtacaatttggttgggtggtggtgc 3458
|||||
Db 125 ATTTGCTCACTCTCTGACAGCAAGCCAGTACGTACAATTTGTTGGTGGTGGTGGTGC 184
QY 3459 tccaaagccacgctcctctgaattgattttttgaattttgattttgnttgnagaagatcacaca 3518
|||||
Db 185 TCCAAGGCCACGCTGCTCTGAAATGATTTTTTGAAGTTT -GTTTGAAGATGATCACA 242
QY 3519 gncatgttacactgactctnaagggacatatntlataacccttttaaaaaaaatccctgt 3578
|||||
Db 243 GTCATGTTACACTGATCTAAAGG--ACATATATATAACCCCTTTAAAAAATACTACTG 299
QY 3579 cctcattcttatttcgagatgaatttcogatacagactagatgcttctctgaagatcaatt 3638
|||||
Db 300 CCTCATTTCTTATTTCAAGATGAATTTCTATACAGACTAGATGTTTTTCTGAAGATCAATT 359
QY 3639 agacattntgaaaatgaatttaaaagtgttttcccttaattgctctctgaaacaagtctctt 3698
|||||
Db 360 AGACATTTTGGAAATGATTTAAAGTGTTTTCCCTTAATGTTCTCTGAAAAACAAGTTTC-TT 418
QY 3699 tgagttttaaccacaaaagtgccctttttgtcacgtggtttctctctagcattctcatgatt 3758
|||||
Db 419 TGTAGTTTTTAACCAAAAAGTGGCC-TTTTGTCACTGGATTCTCCTAGCATTCATGATTT 477
QY 3759 tttttccacaatgaatttaaaattgcttaaaatcagtgagctggtctctctggttgatttc 3818
|||||
Db 478 TTTTTCATACAATGAATTAATAATTCGTAATAATCATGGACTGGCTTTCTGTTGGATTTTC 537
QY 3819 agctaagatgtgtttaagccagagcttttctcagattgtattttttccccaatttt 3878
|||||
Db 538 AGTAAGATGTGTTTAAGGCCAGAGCTTTTCTCAGTATTGATTTTTTTTCCCAATATTT 597
QY 3879 gatttttataacacacacacacacacacacacacacacacacacacacacacacacacac 3938
|||||
Db 598 GATTTTAAAAAATATACACATAGGTGCTGCAATT-ATATCTGCTGCTTTAAATTTCTGTC 656
QY 3939 anatttcactcttagccttttagtagtggnnaatcaanaatttacttttacttaagcatttg 3998
|||||
Db 657 ATATTTCACCTCTAGGCTTTTAGTAGTGGCAATCATATCTTACTTTTACTTAAGCATTT 716
QY 3999 taatttggagtatctgctactagctaagaaataatcnaataattgattttgtactcnc 4058
|||||
Db 717 GTAATTGGAGTATCTGCTAGCTAGGCTAAGAAATTAATCATATAAATGGAG--TTGTACTACC 774
QY 4059 aanatgggtcattctctcatg 4079
|||||
Db 775 ATATATGGATCATTTCCCCAGG 795

RESULT 14
BF608004 769 bp mRNA EST 01-APR-2001
LOCUS MYL1_000946 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
DEFINITION CDNA clone ICRFp522M2029 5', mRNA sequence.
ACCESSION BF608004
VERSION BF608004.1 GI:13504496
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Yahyawi,M., Hennig,S., Neidhardt,L., Radelof,U., Hermann,B.G.,
Lehrach,H. and O'Brien,J.
JOURNAL Detection of a high number of novel genes in a 9-day mouse embryo
cDNA library normalised by oligonucleotide fingerprinting
Unpublished (2001)
```

COMMENT

Contact: Hennig S
Laboraty 123, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
EST's are made from clones being representatives of clone clusters.
Clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-CAGCTATTCAGAGTACTGA-3'
BACKWARD: 5'-TAATGAGCTACTATAGG-3'
Seq primer: 5'-ATTGAGTGACACTATAG-3'
High quality sequence stop: 769.
Location/Qualifiers
1. .769
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ICRFp522M2029"
/dev_stage="embryonic 9-day"
/lab_host="E.coli, XL1 blue"
/note="Vector: PVSportII; Site_1: NotI; Site_2: SalI;
Library preparation by oligo-dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."
208 a 195 c 191 g 174 t 1 others

FEATURES

source

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BASE COUNT

ORIGIN

Query Match 14.8%; Score 617; DB 11; Length 769;
Best Local Similarity 91.5%; Pred. No. 1.1e-88;
Matches 664; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 1195 gcagaatgacgtgtatataatccagaacgcactattacagttaaggaatgttgagac 1254
|||||
Db 44 GCAGGAACGTGCGTGTACATCCGGAACGCCACCATTCAGTGGAAGGCGTTTGAGAC 103
Qy 1255 atgtccaaagttagggagatcataagaanaacatcaggagctttatgaatgatat 1314
|||||
Db 104 GTGTCCCAAGCGGAGGAGGAATAATGAAGAAGATCAGGGAGTCTTATGAAATGATAT 163
Qy 1315 tgccttatgaatctcagcacattatctctgattataatctgaacgccttgggtct 1374
|||||
Db 164 TGCTTCCATGATCTTCAGGACATTTAATCCCTGGATTAAATCTGATGCTTGGGTCT 223
Qy 1375 gtccaccacacttcaggatgacacccctccacactcagggccccccttcagccactcc 1434
|||||
Db 224 GTTCCACCCACGTGAGGATGCCACCTCCACCTCAGGGCCCTTCAACCTGACTCC 283
Qy 1435 tccctaccgcagtttgagcaatacagaacggagactgttccatcagtttccatccagctct 1494
|||||
Db 284 TCCCTTACCACCAATTTGAGCAATCAGACGGAGACTGTGATCTGTTTATTCGCGCCT 343
Qy 1495 atcagtggtgcatcatcgcgaacgagggccagcacatcagcagcttctcgtttgc 1554
|||||
Db 344 GTCCCTTGCGCCATCATTTGGCAACGAGGGCCCAACATCAACAGCTTCTCGCTTTCG 403
Qy 1555 tggagcttcaatgaatgttctccagcggaagcaccagatgctaaagtggagtggtgat 1614
|||||
Db 404 GGGAGCTCGATTAAAGATCGCTCCAGCGGAGCAGCAGATGCTAAAGTGCAGTGTGAT 463
Qy 1615 tatcaatggaccacagaggtcagttcaaggtcaggaagaaattatgaaataataa 1674
|||||
Db 464 TATCACTGGACCACAGAGGTCAGTTCAAGGCTCAGGGAAGAAATTTATGAAAAATTA 523
Qy 1675 aagaacaaacttcttagtctcaagagaggtgaaactgaagctcatatcagagtcc 1734
|||||
Db 524 AGAAGAACACTTTGTTAGTCTTAAAGAGAGGTGAACCTTGAAGCTCAGAGTGCC 583
Qy 1735 atcctttctgctgcagagttattggaagagggcgaacacggtgaatgaacttcagaa 1794
|||||

Db 584 GTCCTTTGCTGCTGGCAGAGTATTGTTGGGAAAGGAGGCAAAATGTGTAATGAGTCCAGAG 643
Qy 1795 ttgtcaagtgcagaagttgtgtccctctgtaccagacacactgatgagaatgaccaagt 1854
|||||
Db 644 TTTATCACGTGCTTGAGTGTGCTCCCGCTGACACACACCTGATGAGAATCATCAAGT 703
Qy 1855 ggtgtcaaaataaactggtcactctatgc-ttgcaggttccccagagagaaatcagg 1913
|||||
Db 704 AGTTGTCAACATAACTGGCCACTTCTATGCTTTGCGGGTCCCGAGAGAAATTCAGG 763
Qy 1914 aaattc 1919
|||||
Db 764 AAATTC 769

LOCUS BG499373 633 bp mRNA EST 27-MAR-2001
DEFINITION 602547471F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4670008 5',
mRNA sequence.
ACCESSION BG499373
VERSION BG499373.1 GI:13460890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 633)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-riemail.nih.gov
Tissue Procurement: DCTP/DRP
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10C1482 row: m column: 17
High quality sequence stop: 618.

FEATURES

source

1. .633
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4670008"
/clone.lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: Sfil (ggcgctcgccc); Site_2: Sfil (ggcattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGCGCATATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGGCGCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

BASE COUNT 219 a 135 c 145 g 134 t

ORIGIN

Query Match 14.6%; Score 610.6; DB 11; Length 633;
Best Local Similarity 98.6%; Pred. No. 1.3e-87;
Matches 616; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1616 atcactggaccaccagaggtcagttcaggtcaggaagaatttatgaaaaataa 1675
|||||
Db 1 ATCACTGGACCACAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGAAAAATTA 60

OM of: US-09-685-696-176 to: N_Geneseq_1101.* out_format : pfs
Date: Jan 25, 2002 11:25 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09685696/runat_25012002_145551_10682/app_query.fasta_1.648
-DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09685696_CGNI_1_331 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-685-696-176
Query length: 579
Database: N_Geneseq_1101.*
Database sequences: 930621
Database length: 428662619
Search time (sec): 163.950000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len | ! Documentation |
|--|-----------|---------|---------|------|-----------------|
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 2956.00 | 4242.42 | 4159 | 3.1e-228 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 2956.00 | 4242.37 | 4181 | 3.2e-228 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 2943.00 | 4232.43 | 4170 | 1.1e-227 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 2187.00 | 3138.19 | 2224 | 1.0e-166 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 1950.50 | 2792.34 | 3412 | 1.9e-147 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 1873.00 | 2686.08 | 1946 | 1.5e-141 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 1868.00 | 2673.59 | 3283 | 7.6e-141 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 1637.00 | 2346.58 | 1708 | 1.3e-122 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 702.50 | 1005.89 | 710 | 5.9e-48 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 640.50 | 921.09 | 444 | 3.1e-43 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 480.00 | 686.47 | 588 | 3.7e-30 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 474.00 | 685.22 | 282 | 4.3e-30 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 341.00 | 490.57 | 364 | 3.0e-19 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 319.50 | 457.99 | 424 | 2.0e-17 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 284.50 | 410.93 | 300 | 8.2e-15 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 241.00 | 327.21 | 2384 | 3.8e-10 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 231.50 | 329.24 | 500 | 2.9e-10 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 231.50 | 318.71 | 1420 | 1.1e-09 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 229.50 | 313.42 | 1803 | 2.2e-09 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 228.50 | 309.17 | 2381 | 3.8e-09 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 227.00 | 335.87 | 136 | 1.2e-10 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 227.00 | 303.65 | 3319 | 7.0e-09 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 219.50 | 296.11 | 2396 | 2.0e-08 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 217.50 | 295.00 | 2007 | 2.3e-08 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 213.50 | 286.66 | 2588 | 6.8e-08 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 204.50 | 273.85 | 2541 | 3.5e-07 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 203.50 | 267.98 | 4219 | 8.2e-07 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 203.50 | 266.98 | 4353 | 8.6e-07 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 192.00 | 268.90 | 2901 | 6.7e-07 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 191.50 | 261.40 | 1357 | 1.7e-06 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 190.00 | 259.00 | 1389 | 2.4e-06 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 187.00 | 251.17 | 1392 | 2.4e-06 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 187.00 | 251.17 | 1966 | 6.5e-06 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 183.00 | 243.37 | 2401 | 1.8e-05 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 183.00 | 240.92 | 3061 | 2.4e-05 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 172.00 | 230.45 | 1790 | 9.3e-05 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 170.50 | 226.72 | 2090 | 0.0001 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 165.50 | 230.35 | 713 | 9.4e-05 |
| /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150 | .. | 159.50 | 214.76 | 1416 | 0.0007 |

/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAH14076 + 158.00 208.68 0.0015
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV41260 + 157.50 204.83 0.0025
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ46281 + 154.50 204.47 0.0026
/SID88/gcgdata/geneseq/geneseq/NA1992.DAT:AAQ30801 + 154.00 200.11 0.0045
/SID88/gcgdata/geneseq/geneseq/NA1996.DAT:AAZ30564 + 153.50 204.96 0.0024

seq_name: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150

seq_documentation_block:

ID AAZ36150 standard; DNA; 4159 BP.
AC AAZ36150;
DT 11-FEB-2000 (first entry)
DE DNA encoding cancer associated antigen KOC-1.
KW Cancer associated antigen; KOC-1; cancer; vaccine; CT7; ss.
OS Homo sapiens.
PN W09954738-Al.
XX 28-OCT-1999.
PF 16-MAR-1999; 99WO-US05766.
PR 17-APR-1998; 98US-0061709.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
DR WPI; 2000-013284/01.
XX Nucleotides representing cancer-associated genes, used to develop
PT products for the diagnosis, monitoring and treatment of cancers
PS Claim 88; Page 39-40; 44pp; English.
CC The present sequence represents a cancer associated antigen gene
CC designated KOC-1. The specification also describes a cancer associated
CC antigen designated CT7. The CT7 polypeptide was isolated from
CC SK-MEL-37 melanoma cells. The polypeptide has some homology with
CC MAGS-10, limited to about 210 carboxy terminal amino acids. The amino
CC terminal of the protein has a repetitive pattern, with repeats rich in
CC serine, proline, glutamine and leucine, and an almost invariable core of
CC the peptide given in AA143877. The CT7 polypeptide can be processed to
CC peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC and polypeptides can be used for treating a cancerous condition and
CC screening for or diagnosing cancerous conditions. The cancer associated
CC antigens can be used as an immunogenic or vaccine composition with an
CC adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
CC stimulating factor (GM-CSF).

Sequence 4159 BP; 1281 A; 830 C; 851 G; 1181 T; 16 other;

alignment_scores:
Quality: 2956.00 Length: 579
Ratio: 5.105 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x AAZ36150 ..
Align seg 1/1 to: AAZ36150 from: 1 to: 4159
1 MetAsn11yLsuyr1leGlyAsnLeuserGluAsnAlaAlaproSerAs 17
|||||
251 ATGAACAACAACTGTATATCGGAACCTCAGGAGAACGCCGCCCTCGGA 300
|||||
17 pLeuGluSerIlePheLysAspAlaIleProValSerGlyProPheL 34

```

|||||
301 CCTAGAAAGTATCTTCAGGAGGCCAAGATCCCGGTGTCGGGACCCCTCC 350
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
351 TGGTGAAGACTGGCTACGCGTTCTGGGACTGCCCGACGAGAGCTGGGCC 400
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
401 CTCGAAGGCCATCGAGGCCCTTTCAGGTAAAATAGAACTGCACGGGAAACC 450
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
451 CATAGAAGTTGAGCACTCGGTCCCAAAAAGGCAAGGATTTCGGAACATTC 500
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
501 AGATACGAAATATCCGCTCATTTACAGTGGGAGTCTGGATAGTTTA 550
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
601 GGAACACTCAGTTGTAATGTAACTATTCCAGTAAGGACCAAGCTAGAC 650
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
651 AAGCACTAGACAAAGTGAATGGATTTACGTTAGAGAAATTTACCTTGAA 700
151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnG 167
701 GTAGCCTATATCCCTGTATGAAATGGCCGCCAGCAAAAACCCCTTGCAGCA 750
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
751 GCCCGGAGTCCCGGGGGCTTGGCAGAGGGGCTCTCAAGGCGAGGGGT 800
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
801 CTCAGGATCCGTATCCAAGCAGAAACCATGTGATTTGCCTCTCGGCGTG 850
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
851 CTGGTCTCCACCCCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 900
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
901 CATTCGGNACATACCAACAGACCCAGTCTAAATCGATCTCCACCGTA 950
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
951 AAGAAATCGGGGCTCTCGAGAAGTCGATTACTATCTCTCTACTCCT 1000
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGl 267
1001 GAAGGCACCTCTCGCGGCTGTAAAGTCTATTCTGGAGATTATGCATAAGGA 1050
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
1051 AGCTCAAGATATAAATTCAGAGAGAGATCCCTTGAAGATTTTAGCTC 1100
284 iAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
1101 ATAATACTTTGTTGGAGCTCTATTGTTAAAGAGGAGGAGAAATCTTAA 1150
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
1151 AAAATTCAGCAAGACACAGACACTTAAATACGATATCTCCATTTGCAGGA 1200
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||

```

.. 4

```

1201 ATTGACGCTGTATATCCAGAACGCACCTATTACAGTTAAAGCAATGTTG 1250
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1251 AGACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATCAGGGAGTCT 1300
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367
|||||
1301 TATGAAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAAATTCCTGG 1350
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
|||||
1351 ATTAATCTGAACGCTTGGGTCTGTTCCACCACCTTCAGGGATGCCAC 1400
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1401 CTCACACCTCAGGGCCCTTTCAGCCATGACTCTCCCTACCCGCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1451 GAGCAATCAGAAACGGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisLysGlnLeuSerArgP 434
|||||
1501 CGGTGCCATCATCGCAAGCAGGCCAGCACATCAAGCAGCTTCTCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1551 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGGAGCACCAGATGCTAAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaGl 467
|||||
1601 GTGAGGATGGTGATTATCCTGCGACCAGCAGAGGCTCAGTTCAAGGCTCA 1650
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
1651 GGGAGGAATTTATGAAAAATTAAGAAAGAAACTTTGTTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1701 AAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTTTGCTGCTGGC 1750
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1751 AGAGTTATTGGAAGAGGAGGCAAAACGGTGAATGAACCTCAGAATTTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1801 AAGTGCAGAAAGTTGTTGCTCTGTCGACACACACCTGATGAGAAATGACC 1850
534 InValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1851 AAGTGGTTGTCAAAATTAACGTGCTCATTCTATGCTGCCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLy 567
1901 AGAAAAATTCAGGAAATTTCTGACTCAGTAAAGCAGCACCACCAACAGAA 1950
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1951 GGCTCTCAAAAGTGGACACCTCAGTCAAGACGGAAG 1987
seq_name: /SDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC65900
seq_documentation_block:
ID AAC65900 standard; cDNA; 4181 BP.
XX
AC AAC65900;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated cDNA L523S;
XX

```

KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
 KW vaccine; detection; ss.

XX Homo sapiens.

OS WO200061612-A2.

XX 19-OCT-2000.

XX 03-APR-2000; 2000WO-US08896.

XX 02-APR-1999; 99US-0285479.

PR 17-DEC-1999; 99US-0466396.

PR 30-DEC-1999; 99US-0476496.

PR 10-JAN-2000; 2000US-0480884.

PR 22-FEB-2000; 2000US-0510376.

XX (CORI-) CORIXA CORP.

PI Wang T, Fan L;

XX WPI; 2000-628399/60.

DR P-FSDB; AAB11328.

XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
 PT protein is used for detecting and monitoring progression of lung cancer
 PT in a patient -

XX Claim 1a; Page 184-186; 261pp; English.

XX This invention describes a novel isolated polypeptide (I) which
 CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
 CC which have cytostatic activity. The polypeptides and polynucleotides are
 CC used in compositions and vaccines to inhibit the development of cancer,
 CC especially lung cancer, in a patient. Methods described in the invention
 CC can be used to monitor the progression of a cancer by carrying out the
 CC detection at subsequent time points and comparing the results from the
 CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
 CC are treated with P2, polynucleotides encoding P2 or antigen presenting
 CC cells expressing P2 and then administered to the patient to inhibit
 CC development of cancer.

XX Sequence 4181 BP; 1303 A; 830 C; 851 G; 1181 T; 16 other;

alignment_scores:

Quality: 2956.00 Length: 579
 Ratio: 5.105 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x AAC65900 ..

Align seg 1/1 to: AAC65900 from: 1 to: 4181

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAalaProSerAs 17

251 ATGAACAACTGTATATCGAACCCTCAGGAGAACGCCGCCCTCGGA 300

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34

301 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGCGGGACCCCTTC 350

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50

351 TGGTGAAGACTGGCTACGGGTTCGTGGACTGCCCGGACGAGACTGGGCC 400

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67

401 CTCAGGCCATCGAGCGCTTCAGGTAATAATAGAACTGCACGGGAACC 450

67 olleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84

1351 ATTAATCTGAACGCCTGGGTCTGTTCACCCACCCACTTCAGGGATGCCAC 1400

451 CATAGAAGTTGAGCACTCGGTCCCAAAAAAGGCAAGGATTTCGGAAACTTC 500

84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100

501 AGATACGAATATCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 550

101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117

551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600

117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134

601 GGAACCTGCAGTTGTAATGTAACTATTCCAGTAAGGACCAAGCTAGAC 650

134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150

651 AAGCACTACACAACTGAATGGATTTCAGTTAGAGAAATTTACACCTTGA 700

151 ValAlaTyrIleProAspGluMetalAlaGlnGlnAsnProLeuGlnGl 167

701 GTAGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGCAGCA 750

167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184

751 GCCCCAGAGTCGCCGGGGCTTGGGCAGAGGGGCTCTCAAGCAGGGGT 800

184 exProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200

801 CTCAGGATCCGTATCCAGCAGAAACCATGTGATTTGCCTCTGCCGCTG 850

201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217

851 CTGGTTCCACCCCAATTTGTTGGAGCATCATAGGAAAGCAAGGTGCCAC 900

217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234

901 CAPTCGGAACATCACCACCAACAGACCAGTCATAAATCGATGTCCACCGTA 950

234 ySgLuAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250

951 AAGAAATCGGGGGCTGCTGAGAAGTCGATTACTATCTCTCTACTCCT 1000

251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMethIlysGly 267

1001 GAAGGCACCTCTCGCGCTTGTAAAGTCTATTCTGGAGATTATGCATAAGGA 1050

267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284

1051 AGCTCAAGATATAAAATTCACAGAAGAGATCCCTTTGAAGATTTTAGCTC 1100

284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300

1101 ATAAATACCTTTGTTGGACGCTTATTGGTAAAGAAAGAAATCTTAAA 1150

301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGl 317

1151 AAAATTTGAGCAAGACACAGACACTAAATCAGCATATCTCCATTGCAGGA 1200

317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334

1201 ATTGACGCTGTATAATCCAGAACGCACATATTACAGTTTAAAGGCAATGTTG 1250

334 luThrCysAlaLysAlaGluGluGluIleMetLysLysIleArgGluSer 350

1251 AGACATGTGCCAAGCTTGAGGAGGAGATCATGAAGAAATACAGGAGTCT 1300

351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367

1301 TATGAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATCTCTGG 1350

367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384

1351 ATTAATCTGAACGCCTGGGTCTGTTCACCCACCCACTTCAGGGATGCCAC 1400

384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1401 CTCACCTCAGGCGCCCTTCAGCCATGACTCCTCCCTACCCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1451 GAGCAATCAGAAACGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgp 434
CGGTGCCATCATCGCAAGCAGGCGCAGACATCAAGCAGCTTCTCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1551 TTGCTGGAGCTTCAATTAAAGATGCTCCAGCGGAAGCACCATGCTATAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG1 467
|||||
1601 GTGAGGATGGTATTCATTCACACGACAGGCTCAGTTCAAGGCTCA 1650
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1651 GCGAAGATTTTATGAAAAATTAAAGAAAGAAACTTTGTTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1701 AAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTTTGCTGCTGGC 1750
501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1751 AGAGTTATGGAAAGGAGGCAAAACGGTGAATGAACCTTCAAGATTTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1801 AAGTGCAGAACTTCTGTCTCCCTCGTACCAGACACCTCATGAGATGACC 1850
534 lnValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1851 AAGTGGTGTGCAAAATAACTGGTCACATCTATGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLy 567
|||||
1901 AGAAAAATTCAGGAATTCAGTCTCAGTAAAGCAGCACCACCAACACAGAA 1950
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1951 GGCTCTGCAAGTGCACACCTCAGTCAAGACGGAAG 1987

seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT:AAC66035

seq_documentation_block:

ID AAC66035 standard; cDNA; 1740 BP.

XX AAC66035;

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

21-FEB-2001 (first entry)

Human lung cancer-associated cDNA antigen L523S.

Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;

vaccine; detection; ss.

OS Homo sapiens.

XX WO200061612-A2.

XX PN

XX PD

XX PF

XX PR

XX PR

XX PR

XX PR

XX PR

XX PR

XX PR

XX PR

XX PR

XX PR

PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX (CORI-) CORIXA CORP.
XX Wang T, Fan L;
XX WPI; 2000-628399/60.
XX P-PSDB; AAB11365.
XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
XX protein is used for detecting and monitoring progression of lung cancer
XX in a patient -
XX
XX Claim 1a; Page 258-259; 261pp; English.
XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
XX are treated with P2, polynucleotides encoding P2 or antigen presenting
XX cells expressing P2, and then administered to the patient to inhibit
XX development of cancer.
XX
XX Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 other;

alignment_scores:

Quality: 2943.00 Length: 579

Ratio: 5.101 Gaps: 0

Percent Similarity: 99.655 Percent Identity: 99.655

alignment_block:

US-09-685-696-176 x AAC66035 ..

Align seg 1/1 to: AAC66035 from: 1 to: 1740

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaProSerAs 17
|||||
1 ATGAACAACTGTATATCGGAACCTCAGCGAGAACGCGCCCTCGGA 50
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
51 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGTCGGGACCCCTTC 100
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
101 TGGTGAAGACTGGCTACGCTTCGTGGACTGCCCGGAGAGAGCTGGGCC 150
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
151 CTCAGGCCCATCGAGGCGCTTTCAGGTAAATAGAACTGCACGGGAACC 200
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
201 CATAGAAGTTGAGCAGCTCGGTCCCAAAAGCAAGGATTCGGAAGACTTC 250
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
251 AGATACGAAATATCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTA 300
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
301 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 350
117 rGluThrAlaValValAsnValThrTyrSerLysAspGlnAlaArgG 134
|||||
351 GGAACATGCAGTTGTAANTGTACCTATTCCAGTAGGACCAAGCTAGAC 400


```

134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
401 AAGCACTAGACAACTGAATGGATTTCAGTTAGAGAATTCACCTTGAAA 450
|||||
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
451 GTAGCCATATCCCTGATGAACGGCGCCAGCAAAACCCCTTGACGCA 500
|||||
167 nProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
|||||
501 GCGCCGAGTCCGGGGGCTTGGCGAGAGGGCTCCTCAAGCGAGGGGT 550
|||||
184 exProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||
551 CTCAGGATCCGTATCCAAGCAGAAACCATGTGATTTGCCTCGCGCTG 600
|||||
201 LeuValProThrGlnPheValGlyAlaIlelleGlyLysGluGlyAlaTh 217
|||||
601 CTGGTTCCACCCCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 650
|||||
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
|||||
651 CATTCGACATCACCAACAGACCCAGTCTAAATCGATGTCACCGTA 700
|||||
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
|||||
701 AAGAAATCGGGGCTGCTGAGAGTCGATTACTATCTCTACTCTCT 750
|||||
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMethLysG 267
|||||
751 GAAGGCACCTCTCGCGCTTGTAACTATCTGGAGATTATGCATAAGGA 800
|||||
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
|||||
801 AGCTCAAGAATATAAAATCCAGAGAGATCCCTTGAAGATTTAGCTC 850
|||||
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
851 ATAATAACTTTGTTGGACGCTTATTGGTAAAGAGAAATCTTAA 900
|||||
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
|||||
901 AAAATTGAGACAGACACAGACATAAAATCAGATATCTCCATTGCAGGA 950
|||||
317 uLeuThrLeuTyrAspProGluArgThrIleThrValLysGlyAsnValG 334
|||||
951 ATTGACGCTGTATAATCCAGAACGCACATTACAGTTAAAGCAATGTTG 1000
|||||
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1001 AGACATGTCGCAAGCTGAGGAGGAGATCATGAGAAATCAGGGAGTCT 1050
|||||
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG 367
|||||
1051 TATGAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATTCCTG 1100
|||||
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetProp 384
|||||
1101 ATTAATATGAACGCCCTGGGTCTGTTCGCCACCCACTTCAGGGATGCCAC 1150
|||||
384 rProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1151 CTCACCACTCAGGGCCCTTTCAGCCATCAGTCTCTCCCTACCGCAGTTT 1200
|||||
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1201 GAGCAATCAGAAACGGAGAGCTGTTTCATCTGTTATCCAGCTCTATCAGT 1250
|||||
417 lGlyAlaIlelleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgP 434
|||||
1251 CGGTGCCATCATCGGCAAGCAGGGCCAGCACATCAAGCAGCTTCTCGCT 1300
|||||

```

```

434 heaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1301 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGGAAGCACAGATGCTAAA 1350
|||||
451 ValArgMetValIlelleThrGlyProProGluAlaGlnPheLysAlaG 467
|||||
1351 GTAGAGATGGTGATTATCACTGGACCACCAGAGGCTCAGTTCAAGGCTCA 1400
|||||
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1401 GCGAAGATTTATGGAATAATTAAGAGAGAAACTTTGTTAGTCTCTAAG 1450
|||||
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1451 AAGAGGTGAAACTTGAAGCTCATATCAGAGTGCATCTCTTTGCTGTGTC 1500
|||||
501 ArgValIlelleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1501 AGAGTTATTGGAAAGAGGCAAAACGGTGAATCACTTCAGAAATTTGTC 1550
|||||
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1551 AAGTGCAGAACTTGTGCTCTCGTGCACAGACACCTGATGAGATGACC 1600
|||||
534 lnValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1601 AAGTGGTTGTCAAAATAACTGGTCACTTCTATGCTTGCAGGTTGCCAG 1650
|||||
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLys 567
|||||
1651 AGAAAAATTCAGGAAATTCAGTACTAGTAAAGCAGCACCACCAACAGAA 1700
|||||
567 sAlaLeuGlnSerGlyProGlnSerArgArgLys 579
|||||
1701 GGCTCTGCAAAAGTGGACCACCTCAGTCAAGACCGGAAG 1737
|||||

```

seq_name: /SID88/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ10617

seq_documentation_block:

ID AAZ10617 standard; cDNA; 2224 BP.

XX AAZ10617;

XX 17-NOV-1999 (first entry)

XX cDNA encoding a murine c-myc coding region determinant binding protein.

KW c-myc coding region determinant binding protein; CRD-BP; tumor;

KW c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer;

KW pancreatic cancer; ss.

OS Mus musculus.

XX Key Location/Qualifiers

FT CDS 131..1854

FT /*tag= a

XX WO9946594-A2.

XX 16-SEP-1999.

XX 05-MAR-1999; 99WO-US04897.

XX 09-MAR-1998; 98US-0077372.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Ross J;

XX WPI; 1999-551506/46.

DR P-PSDB; AAY30649.

XX Diagnosing presence or absence of a tumor in a human by examining c-myc

PT coding region determinant-binding protein

XX Example; Fig 1A-D; 79pp; English.

XX The present sequence encodes a murine c-myc coding region determinant binding protein (CRD-BP). The presence or absence of a tumor can be determined by determining the levels of CRD-BP present in the suspect tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack and so prolongs its half-life. The methods are used for diagnosing CC presence or absence of a tumor in a human, especially breast, colon CC and pancreatic cancer. They are also used to inhibit cancer cell growth.

XX Sequence 2224 BP; 580 A; 609 C; 618 G; 417 T; 0 other;

alignment_scores:

Quality: 2187.00 Length: 585
Ratio: 4.150 Gaps: 8
Percent Similarity: 90.085 Percent Identity: 74.188

alignment_block:

US-09-685-696-176 x AAZ10617 ..

Align seg 1/1 to: AAZ10617 from: 1 to: 2224

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
131 ATGAACAAGCTTTACATCGGCAACCTTCAACGAGAGTGTGACCCCGCAGA 180
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
181 CTTGGAGAAAGTATTCTCGGAGCACAAGATCTCTACAGCGCGCGAGTTCT 230
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
231 TGGTCAATCGGGTACGCCCTTCGTGGATGCCCCGACGACACTGGGCG 280
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
281 ATGAAGGCCATCGAAACTTTCTCGGGGAAAGTAGAACTGCAAGGAAACG 330
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
331 TCTAGAGATTGAACACACTCAGTCCCAAAACAAAGAGGTCGGAATAAC 380
84 InIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
381 AGATCCGCAATATTCACCTCAGCTCCGATGGGAAGTGTAGATAGCTG 430
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
431 CTGGCTCAGTACGGTACAGTGGAGAACTGTGAGCAAGTGAACACTGAAAG 480
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
481 TGAGACAGCGGTGGTCAACGCTCACTACTTAACCCGGGAGCAGACCAGGC 530
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
531 AAGCTATCATGCAAGCTAAATGGCCATCACTGGAGAACCATGCCCTGAAG 580
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
581 GTCTCTACATACCTGATGACGATAACA...CAAGGTCCTGAGAAATGG 627
167 nProArgLysArgArgGlyLeuGlnArgGlySerSerArgGlnGlyS 184
628 GCGTCGTGGA.....GGCTTTGGGTCTCGGGGCCAGCCCGCAAGGGT 671
184 erPro.....GlySerValSerGlnLysProCysAspLeuPro 197
672 CGCCCGTGGCAGCAGGGGCTCCAGCCCAAGCAGCAGCAGTGGACATCCCT 721

198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGl 214
722 CTCGGGCTCCTGGTGGCTACGCAGTATGTAGCGCTATCATTTGCAAGGA 771
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
772 GGGTGGCCACCATCGCAACATCACAAACACAGCAGCTCCAAATAGACG 821
231 alHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
822 TGCATAGAGAGAGATCGGGCGCTGCGGAGAGAGGCCATCAGCGTGCAT 871
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMe 264
872 TCAACCCCTGAAGGCTGCTCCTCGCGTGCAGATGATCTTTGGAGATTAT 921
264 tHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
922 GCACAAGGAGGCAAGGACACCAAAACGGCAGATGAAGTTCCCTGAAGA 971
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
972 TCCTGGCTCATAACAACCTTCGTCGGCGACTCATTTGGCAAGAGGCCGG 1021
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
1022 AACCTGAAGAAGGTGGAGCAGCAGACAGACGAAGATCACCATCTCATC 1071
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
1072 GCTCCAGGACCTCACGCTCTATAACCTCGAGAGGACCATCACTGTGAAG 1121
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
1122 CGGCATTTGAACTGTTGCGAGGCCCGCAGCAGGAGATCATGAAGAAGTT 1171
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
1172 CGAGAGCTTACGAGAACGACGAGTGGCGCCCATGAGTTGCAGTCCCACCT 1221
364 uIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerG 381
1222 CATCCCTGGGCTTAACCTGGCTGCTGAGGTCTCTCCACGCTTCAATCCA 1271
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
1272 GCGTGTCCCTCCT.....CCTCCAGCAGTGTCACTGGGGCTGCT 1312
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
1313 CCTTATAGTCTCTCATGCGAGCTCCGCGAGCAGGAGATGGTACAAGTGT 1362
411 eIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlnHisI 428
1363 CATCCCGCCCGAGGCTGGGCGCCATCATTTGGCAAGAAGGGCCAGACA 1412
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
1413 TCAACAACCTCTCCCGTTTCCGACGCGCTCCATCAAGATTGCTCCACCA 1462
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProGl 461
1463 GAAACACCTGACTCCAAAGTTGGAATGGTCTCATCATCTACTGACCCCCAGA 1512
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluAla 478
1513 GGTCTAGTTCAAGGCTCAGGGAAGATTTATGGCAAACTAAAGAAGAGA 1562
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
1563 ATTTCTTTGGTCCCAAGAGAGAGTAGAAGCTAGAGACCCACATACGGGTT 1612

250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisIy 266
811 CCAGAGGGGACTTTCGAGCATGCGCGCATGATCTTGAATCATCGCAA 860
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
861 AGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCTCTGAAATCTTGG 910
283 IaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
911 CACACAATGGCTTGGTTGGAAGACTGATTGGAAAAAGAGGAGCAAAATTG 960
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG 316
961 AAAAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 1010
316 nGluLeuThrLeuThrAsnProGluArgThrIleThrValLysGlyAsnV 333
1011 GGATTGAGCATATACACCGGAAAGAACCATCACTGTGAGGCGCACAG 1060
333 alGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArgGlu 349
1061 TTGAGGCTGTGCCAGTGTGATGATAGAGATTATGAAGAAGCTGCGTGAG 1110
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1111 GCCTTTGAAATGATATCTGGCTGTTAAACCAACAAAGCAATCTGATCCC 1160
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe..... 376
1161 AGGTTGAACTTCAGCGCAGCTTGGATCTTTTCAACAGGAGTGTCCGTGC 1210
377ProProThrSer.....GlyMetProProProThr..... 386
1211 TATCTCCACAGCAGCGCGCGGAGCTCCCGCGCTGCCCGCTACCAC 1260
387SerGlyProProSerAlaMetThrPro..... 395
1261 CCCTTCACTACCCACTCCGGATCTTCTCCAGCGCTGTACCCCGCTACCA 1310
396ProTyrProGlnPheGluGln...SerGluThrGluThrValH 409
1311 GTTTGGCCGTTCCCGCATCATCTCTATCCAGCAGCAGGATTGTGA 1360
409 IsGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425
1361 ATCTCTTCATCCCAACCCAGCGCTGTGGCGCATCATCGGAAGAAGGG 1410
426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAl 442
1411 GCACACATCAACAGCTGGGAGATTCCCGGAGCTCTATCAAGATTGC 1460
442 aProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyP 459
1461 CCCTGCGGAAGCCAGACGCTCAGCGAAGGATGGTCATCATCACCGGCG 1510
459 roProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLys 475
1511 CACCGGAAGCCAGTTCAAGCCCGCAGGACGATCTTTGGGAAGCACTGAA 1560
476 GluGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisI 492
1561 GAGGAAATCTCTTTAAACCCCAAGAAAGTGAAGCTGGAAGCGCATAT 1610
492 eArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysT 509
1611 CAGAGTGCCTCTTCCACAGCTGGCGGGGTGATTGGCAAGAGTGCACAGA 1660
509 hrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArg 525
1661 CCGTGAGCAACTGCAGAACTTAACAGTGCAGAAAGTCACTGCTGCTGCT 1710
526 AspGlnThrProAspGluAsnGlnValValValLysIleThrGlyHI 542

1711 GACCAACGCCAGATGAAAATGAGGAAGTGCATGAGAAATATATCGGCA 1760
542 sPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrG 559
1761 CTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAATTTGTACAC 1810
559 InValLysGlnHisGlnGlnGln 1866
1811 AGGTGAAGCAGCAGGAGCAAAA 1833
seq_name: /SID58/gcgdata/geneseq/geneseq/Na2000.DAT:AAZ36153
seq_documentation_block:
ID AAZ36153 standard; DNA; 1946 BP.
XX AAZ36153;
XX (first entry)
DT 11-FEB-2000
XX An alternative form of DNA encoding cancer associated antigen KOC-2.
DE DE
XX Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.
KW KW
XX Homo sapiens.
OS OS
XX WO9954738-A1.
PN PN
XX 28-OCT-1999.
PD PD
XX 16-MAR-1999; 99WO-US05766.
PF PF
XX 17-APR-1998; 98US-0061709.
PR PR
XX (LUDW-) LUDWIG INST CANCER RES.
PA PA
XX Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
PI PI
XX WPI; 2000-013284/01.
DR DR
XX Nucleotides representing cancer-associated genes, used to develop
PT PT products for the diagnosis, monitoring and treatment of cancers -
PT PT
XX Claim 55; Page 42; 44pp; English.
PS PS
XX The present sequence represents an alternative form of a cancer
CC CC associated antigen gene designated KOC-2. The specification also
CC CC describes a cancer associated antigen designated CT7. The CT7
CC CC polynucleotide was isolated from SK-MEL-37 melanoma cells. The
CC CC polypeptide has some homology with MAGP-10, limited to about 210 carboxy
CC CC terminal amino acids. The amino terminal of the protein has a repetitive
CC CC pattern, with repeats rich in serine, proline, glutamine and leucine,
CC CC and an almost invariable core of the peptide given in AAY43877. The CT7
CC CC polypeptide can be processed to peptides which provoke lysis by
CC CC cytolytic T cells. The polynucleotides and polypeptides can be used for
CC CC treating a cancerous condition and screening for or diagnosing cancerous
CC CC conditions. The cancer associated antigens can be used as an immunogenic
CC CC or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or
CC CC granulocyte macrophage-colony stimulating factor (GM-CSF).
XX XX
SQ Sequence 1946 BP; 502 A; 528 C; 553 G; 358 T; 5 other;
alignment_scores:
Quality: 1873.00 Length: 509
Ratio: 4.116 Gaps: 9
Percent Similarity: 89.391 Percent Identity: 74.656
alignment_block:
US-09-685-696-176 x AAZ36153 ..
Align seg 1/1 to: AAZ36153 from: 1 to: 1946


```
187 erValSerLysGlnLysProCysAspLeuProLeuArgLeuValPro 203
    ::::::::::: ::::::::::: |||||
114 GCACCTTCTCAGGCCAGACAGATGATTCGCCGTCGGATCTCGTCCCC 163
    ::::::::::: ::::::::::: |||||

204 ThrGlnPheValGlyAlaIleGlyLysGluGlyAlaThrIleArgAs 220
    |||||
164 ACCCATTTGTTGGTGCCATCATCGGAAGAGGGCTTGACCCATAAGAA 213
    ::::::::::: ::::::::::: |||||

220 nIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnA 237
    |||||
214 CATCACTAAGCAGACACCCAGTCCCGGTAGATATCCATAGAAAGAGAACT 263
    ::::::::::: ::::::::::: |||||

237 laGlyAlaIaIaGluLysSerIleThrIleLeuSerThrProGluGlyThr 253
    |||||
264 CTGGAGCTGCAGAGAAGCCCTGTGCACATCATCCATGCCACCCAGAGGCACT 313
    ::::::::::: ::::::::::: |||||

254 SerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAlaGlnAs 270
    |||||
314 TCTGAAGCATGCCGATGATCTTGAATCATGCAGAAGAGGCAGATGA 363
    ::::::::::: ::::::::::: |||||

270 pIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnP 287
    ::::::::::: ::::::::::: |||||
364 GACCAAACTAGCCGAAGAGATTCTCTGAAATCTTGGCACACAATGGCT 413
    ::::::::::: ::::::::::: |||||

287 heValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysIleGlu 303
    ::::::::::: ::::::::::: |||||
414 TGGTTGGAAGACTGATTTGGAAGAGAGGCAAAATTTGAAGAAATTTGAA 463
    ::::::::::: ::::::::::: |||||

304 GlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuThrLe 320
    ::::::::::: ::::::::::: |||||
464 CATGAACAGGAGGACCAAGATAACAACTCTCATCTTGCAGGATTTGAGCAT 513
    ::::::::::: ::::::::::: |||||

320 uTyRAsnProGluArgThrIleThrValLysGlyAsnValGluThrCysA 337
    |||||
514 ATACAACCCGGAAGAACCATCACTCTGAAGGCACAGTTGAGGCCCTGTG 563
    ::::::::::: ::::::::::: |||||

337 laLysIaIaGluGluIleMetLysLysIleArgGluSerTyRLeuAsn 353
    |||||
564 CCAGTCTGAGATAGAGATTATGAAGAAACTGCGTGAGGCCCTTTGAAAT 613
    ::::::::::: ::::::::::: |||||

354 AspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGly...LeuAs 369
    |||||
614 GATATCTGCTGTAACTGTAAGTCCCTAAATGCTTCTTCNCGCTGGG 663
    ::::::::::: ::::::::::: |||||

369 nLeuAsnAlaLeuGlyLeuPheProThrSer...GlyMetPro 383
    ::::::::::: |||
664 TTTCACTAGGCTAAAAATCTTGCCATTCAGCTNATGAGGAATGCCT 709
    ::::::::::: |||||
```

seq_name: /SID8/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC17226

seq_documentation_block:

ID AAC17226 standard; cDNA; 444 BP.

AC AAC17226;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 21301.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

```
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 21301; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 444 BP; 145 A; 88 C; 91 G; 120 T; 0 other;
```

alignment_scores:

Quality: 640.50 Length: 138

Ratio: 4.744 Gaps: 3

Percent Similarity: 97.826 Percent Identity: 97.101

alignment_block:

US-09-685-696-176 x AAC17226 ..

Align seg 1/1 to: AAC17226 from: 1 to: 444

228 LysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIl 244

::::::::::: |||||

38 AGAATCGATGTCACCCGTAAGAAATGCGGGGCTGCTGAGAAGTCGAT 87

244 eThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleL 261

::::::::::: |||||

88 TACTATCCTCTCTACTCTCTGAAGGCACCTCTGCGGCTTGTAACTTATTC 137

261 euGluIleMetHisLysGluAlaGlnAspIleLysPheThr.GluGluIl 277

::::::::::: |||||

138 TGGAGATTATGCATAAGGAAGCTCAAGATATAAAATTCACAAGAG...AT 184

277 eProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyL 294

::::::::::: |||||

185 CCCCTTGAAGATTTAGCTCATATAACTTTGTTGGACGCTCTTATGGTA 234

294 ySGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIle 310

::::::::::: |||||

235 AAGAAGGAAGAATCTTAAAAAATTTGAGCAAGACACACACATAAAATC 284

311 ThrIleSerProLeuGlnGluLeuThrLeuTyRAsnProGluArgThrIl 327

::::::::::: |||||

285 ACGATATCTCCATTGCGAGGAATTTGACGCTGTATAATCCAGAAGCAGCAT 334

327 eThrValLysGlyAsnValGluThrCys.AlalysAlaGluGluIle 343

::::::::::: |||||

335 TACAGTTAAAGGCAATGTTGAGACATGTTGCCAAAGCTGAGGAGGATC 384

344 MetLysLysIleArgGluSerTyRLeuAsnAspIleAlaSerMetAsnLe 360

::::::::::: |||||

385 ATGAAGAAATCAGGGAGTCTTTATGAAAATGATATTGCTTCTATGAATCT 434

360 uGlnAlaHis 363

::::::::::: |||||

435 TCAAGCACAT 444

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI43398

seq_documentation_block:

ID AAI43398 standard; DNA; 588 BP.

XX AC AAI43398;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #12084 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX OS genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX PS Claim 25; SEQ ID No 12084; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 588 BP; 156 A; 139 C; 143 G; 150 T; 0 other;

alignment_scores:

Quality: 480.00 Length: 95

Ratio: 5.053 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.947

alignment_block:

US-09-685-696-176 x AAI43398 ..

Align seg 1/1 to: AAI43398 from: 1 to: 588

134 GlnAlaLeuAspTysLeuAsnGlyPheGlnLeuGluAsnPhetrLeuLy 150

:::|||||

19 AGAGCACTAGACAAACTGAATGGATTTTCAGTTAGAGAATTTACCTTGAA 68

150 sValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167

|||||

69 AGTAGCCTATATCCTGATGAATGGCCGCCAGCAAAACCCCTTGACG 118

167 lnProArgGlyArgGlyLeuGlyGlnArgGlySerArgGlyGlnGly 183

|||||

119 AGCCCCAGGTGCGCGGGGCTTGGGAGAGAGGGGCTCTCAAGGCAGGG 168

184 SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLe 200

|||||

169 TCTCCAGGATCCGTATCCAGCAGAAACCATGTGATTTGCCTCTCGGCT 218

200 uLeuValProThrGlnPheValGlyAlaIleGlyLysGlyGlyAlaT 217

|||||

219 GCTGGTTCCCAACCAATTTGTTGGAGCCATCATAGAAAAGAGTGCCA 268

217 hrIleArgAsnIleThrLysGlnThrGlnSerLys 228

|||||

269 CCATTCGGAACATCCCAACAGACCCAGTCTAAG 303

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI56378

seq_documentation_block:

ID AAI56378 standard; DNA; 282 BP.

XX AC AAI56378;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #25064 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX OS genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX PS Claim 25; SEQ ID No 25064; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 282 BP; 77 A; 79 C; 70 G; 56 T; 0 other;

alignment_scores:

Quality: 474.00 Length: 93

Ratio: 5.097 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x AAI56378 ..

Align seg 1/1 to: AAI56378 from: 1 to: 282

```
135 AlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysVa 151
|||||
2  GCACTAGACAAACTGAATGGATTTCAGTTAGAGAAATTCACCTTGAAGT 51
|||||
151 lAlaTyrlleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnClnp 168
|||||
52  AGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGCAGCAGC 101
|||||
168 rOArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGlySer 184
|||||
102  CCCGAGGTCCCGGGGCTTGGCAGAGGGGCTCCTCAAGGCAGGGGTCT 151
|||||
185 ProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeuLe 201
|||||
152  CCAGGATCCGTATCCCAAGCAGAAACCATGTGATTTGCCCTCTCGCCCTGCT 201
|||||
201 uValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrI 218
|||||
202  GGTTCACCAACCAATTTGTTGGAGCCATCATAGAAAAGAAAGGTGCCACCA 251
|||||
218 lEArgAsnIleThrLysGlnThrGlnSer 227
|||||
252  TTCGGAACATCACCAAAACAGACCCAGTCT 280
|||||
```

seq_name: /SID8/9c9data/geneseq/geneseqn/NA2000.DAT:AAC03267

seq_documentation_block:

ID AAC03267 standard; cDNA: 364 BP.

AC AAC03267;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 3265.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GBST) GENSET.

PA Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR P-PSDB; AAG03261.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 3265; 7lpp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
mRNAs encoding secreted proteins. An ORF has been identified within the
sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
derived from 30 different tissues. EST sequences usually correspond
mainly to the 3' untranslated region (UTR) of the mRNA because they are
often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
well suited for isolating cDNA sequences derived from the 5' ends of
mRNAs and even in those cases where longer cDNA sequences have been
obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
mRNAs with intact 5' ends and can therefore be used to obtain full length
cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion
CC vectors.

XX Sequence 364 BP; 84 A; 101 C; 118 G; 59 T; 2 other;

alignment_scores:
Quality: 341.00 Length: 96
Ratio: 4.060 Gaps: 0
Percent Similarity: 87.500 Percent Identity: 64.583

alignment_block:

US-09-585-696-176 x AAC03267 ..

Align seg 1/1 to: AAC03267 from: 1 to: 364

1 MetAsnLysLeuTyrlleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||

76 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGCTCACCGCGACGA 125
|||||

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||

126 CCTCCGCGACGCTTTGGGACAGAGAAAGCTGCCCTGGCGGACAGGTCC 175
|||||

34 euValLysThrGlyTyrlleAlaPheValaspCysProaspGluSerTrpAla 50
|||||

176 TGCTGAAGTCCGGCTAGCGCTTCTGGACTACCCGACCAAGAACTGGGCC 225
|||||

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||

226 ATCCGCGCATCGAGACCCCTCTCGGGTAAAGTGAATTCATGGGAAAT 275
|||||

67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||

276 CATGGAGAGTGATTACTCAGTCTCTAAAAGCTAAGGACGAGGAAATTC 325
|||||

84 lnIleArgAsnIleProProHisLeuGlnTrpGluVal 96
|||||

326 AGATTGGAACATCCCTCCTCACCTGCAGTGGGAGGTG 363
|||||

seq_name: /SID8/9c9data/geneseq/geneseqn/NA2000.DAT:AAH31011

seq_documentation_block:

ID AAH31011 standard; cDNA: 424 BP.

AC AAH31011;

XX 27-JUL-2001 (first entry)

DE Human colon cancer cell line Kml2L4-A cDNA library derived sequence #945.

KW Human; diagnosis: colon cancer; cancer; malignant; chromosome mapping;

KW detection; colon cancer cell line Kml2L4-A; ss.

OS Homo sapiens.

PN WO200018916-A2.

XX 06-APR-2000.

XX 23-SEP-1999; 99WO-US22226.

XX 28-SEP-1998; 98US-0102161.

PR 28-SEP-1998; 98US-0102180.

PR 29-SEP-1998; 98US-0102380.

PR 08-OCT-1998; 98US-0103815.

PR 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;


```
|||||
17 CCGATGAGAGGTGAGCTCCCTTCGCCCTCAG.....CGAGCCCA 60
   |||
171 gArgGlyLeuGlyGlnArgGlySerArgGlnGly...SerProGlyS 187
   |||
61 GCGT.....GGGACCATCTCTCCGGGAGCAAGGCCAGCCCTGGGG 104
187 erValSerLysGlnLysProCysAspLeuProLeuArgLeuLeuValPro 203
   |||
105 GCATTCTCAGGCCAGACAGATTGATTCCCGCTCGGATCCTGTCCC 154
204 ThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAs 220
   |||
155 ACCAGTTTGTGGTCCCATCATCGAAAGGAGGGCTTCACCATAAAGAA 204
220 nileThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnA 237
   |||
205 CATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACT 254
237 laGlyAlaAlaGluLysProCysIleThrIleLeuSerThrProGlu 251
   |||
255 CTGGAGCTGCAGAGAAGCCTGTCAACCATCCATGCCACCCACAGAG 298
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 14:55:39 ; Search time 65.7 Seconds
(without alignments)
652.792 Million cell updates/sec

Title: US-09-685-696-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAPSDLES.....VKHQHQKALQSGPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_1101.*
- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 2956 | 100.0 | 579 | 21 AAB11328 | Human lung cancer- |
| 2 | 2943 | 99.6 | 579 | 21 AAB11365 | Human lung cancer |
| 3 | 2190 | 74.1 | 577 | 20 AAY30649 | A murine c-myc cod |
| 4 | 474 | 16.0 | 93 | 22 AAM38501 | Peptide #12538 enc |
| 5 | 341 | 11.5 | 97 | 21 AAG03261 | Human secreted pro |
| 6 | 241 | 8.2 | 644 | 15 AAR58813 | Human c-myc far up |
| 7 | 231.5 | 7.8 | 313 | 21 AAG08900 | Arabidopsis thalia |
| 8 | 229.5 | 7.8 | 590 | 15 AAR58816 | Human c-myc far up |
| 9 | 228.5 | 7.7 | 643 | 15 AAR58814 | Human c-myc far up |
| 10 | 227 | 7.7 | 45 | 22 AAM37315 | Peptide #11352 enc |
| 11 | 227 | 7.7 | 530 | 13 AAR28888 | Ri paraneoplastic |

| | | | | | |
|----|-------|-----|------|-------------|--------------------|
| 12 | 217.5 | 7.4 | 644 | 21 AAG39346 | Arabidopsis thalia |
| 13 | 213.5 | 7.2 | 403 | 21 AAB42547 | Human OREF ORF2311 |
| 14 | 213.5 | 7.2 | 403 | 22 AAB94825 | Human protein sequ |
| 15 | 204.5 | 6.9 | 781 | 21 AAG31213 | Arabidopsis thalia |
| 16 | 204.5 | 6.9 | 803 | 21 AAG31212 | Arabidopsis thalia |
| 17 | 204.5 | 6.9 | 846 | 21 AAG31211 | Arabidopsis thalia |
| 18 | 201.5 | 6.8 | 590 | 21 AAG39347 | Arabidopsis thalia |
| 19 | 201.5 | 6.8 | 1195 | 11 AAR05530 | High density lipop |
| 20 | 201.5 | 6.8 | 1292 | 11 AAR05531 | High density lipop |
| 21 | 191.5 | 6.5 | 419 | 21 AAG21062 | Arabidopsis thalia |
| 22 | 191.5 | 6.5 | 452 | 21 AAG21061 | Arabidopsis thalia |
| 23 | 190.5 | 6.4 | 462 | 22 AAB49335 | Human hnrNPK prote |
| 24 | 190 | 6.4 | 463 | 14 AAR3766 | Rat liver catalase |
| 25 | 190 | 6.4 | 463 | 20 AAY39467 | Heterogenous ribon |
| 26 | 190 | 6.4 | 464 | 14 AAR38906 | Rat liver catalase |
| 27 | 187 | 6.3 | 465 | 21 AAB43753 | Human cancer assoc |
| 28 | 187 | 6.3 | 465 | 22 AAG73800 | Human colon cancer |
| 29 | 181.5 | 6.1 | 446 | 21 AAG39348 | Arabidopsis thalia |
| 30 | 172 | 5.8 | 495 | 21 AAG39951 | Arabidopsis thalia |
| 31 | 172 | 5.8 | 500 | 21 AAG39950 | Arabidopsis thalia |
| 32 | 170.5 | 5.8 | 629 | 21 AAG36715 | Arabidopsis thalia |
| 33 | 167.5 | 5.7 | 228 | 21 AAG08901 | Arabidopsis thalia |
| 34 | 159.5 | 5.4 | 229 | 21 AAG21063 | Arabidopsis thalia |
| 35 | 159.5 | 5.4 | 471 | 22 AAB30930 | Amino acid sequenc |
| 36 | 156.5 | 5.3 | 571 | 21 AAG36716 | Arabidopsis thalia |
| 37 | 154.5 | 5.2 | 639 | 21 AAG40000 | Arabidopsis thalia |
| 38 | 154 | 5.2 | 343 | 13 AAR28887 | pR18 fusion protei |
| 39 | 153.5 | 5.2 | 230 | 17 AAR95954 | Eukaryotic cell gr |
| 40 | 151.5 | 5.1 | 418 | 21 AAG39952 | Arabidopsis thalia |
| 41 | 150 | 5.1 | 623 | 19 AAW68095 | Chlamydomonas rein |
| 42 | 147 | 5.0 | 243 | 15 AAR58815 | Human c-myc far up |
| 43 | 146 | 4.9 | 650 | 21 AAY85180 | Cellulose synthase |
| 44 | 145.5 | 4.9 | 414 | 17 AAW10529 | Saccharomyces cere |
| 45 | 145.5 | 4.9 | 414 | 22 AAB30809 | Amino acid sequenc |

ALIGNMENTS

RESULT 1
AAB11328
ID AAB11328 standard; Protein; 579 AA.
XX
AC AAB11328;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated protein L523S.

XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
vaccine; detection.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX

PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
(CORI-) CORIXA CORP.
PI Wang T, Fan L;
XX
DR WPI; 2000-628399/60.
DR N-PSDB; AAC65900.
XX

PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -
XX
PS Claim 3; Page 186-188; 261pp; English.
XX
XX This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.
XX
SQ Sequence 579 AA;

Query Match 100.0%; Score 2956; DB 21; Length 579;
Best Local Similarity 100.0%; Pred. No. 4.9e-246;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNKLYIGNLSENAPDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKATEALS GK 60
Db 1 mnklyignlsenapsdlesifkdakipvsgpflvktgyafvdcpdsewalkatealsgk 60
QY 61 IELHGKPIEVHSEVPRQRIRKLIQIRNIPPHLQWEVLDSLLOVGVVESCSEQVNTDSETA 120
Db 61 ielhgkpievehsvprqrirklqirnipphlqwevldsllovgvvescseqvntdseta 120
QY 121 VVNVYSSKQDQARQALDKLNGFOLENTFLKVAYIPDDEMAAQNPLOQPRGRRLGQRGSS 180
Db 121 vvnvtysskqdarqaldklngfqlentflkvayipdemaagnploqprgrrlgqrgss 180
QY 181 RQSPGVSVKQKPCDPLRLVPTQFVGALIGEGATIRNITKTOSKIDVHRKENAGAA 240
Db 181 rgspgsvskqkpcdplrlvptqfvgaligegatirnitktoskldvhrkenagaa 240
QY 241 EKSTITLSTPEGTSAACKSILEIMHKEAQDIKFTTEEIPKLAHNNFVGRGLIGKEGRNLK 300
Db 241 eksitiltstpegtsaacksileimhkeaqdikftteeeipklahnnfvgrligrkegrnlk 300
QY 301 KIEQDQDTKITISPLQELTYLNPRTITVKGNVETCAKABEEIMKKIRESYENDIASMNL 360
Db 301 kieqdttdtkitispqlqeltylnpirtitvkgnvetcakabeeimkkiresyendiasmnl 360
QY 361 QAHLIPGLNLAGLPPTSGMPPPTSGPPSAMTPPYQPEQSETEVHQFIPALSVGAI 420
Db 361 qahlipglnlaglpptsgmppptsgppsamtpypyqpeqsetevhqfipalsvgai 420
QY 421 IGKOGQHIKOLSRFAGASIKIAEAPDAKVRVVIITGPPPEAQFKAQGRITGKIKENFV 480
Db 421 igkoghikolsrfagasi kieapdavrvmvviitgpppeaqfkaqgritgkikenfv 480
QY 481 SPKEEVKLEAHIRVPFAACRGVKGKGTNQLNLSAEVVPVPRDQTPDENQDVVVKIT 540
Db 481 spkeevkleahirvpfaacrgvkgkgtvnqlnlsaevvppvprdqtpdenqdvvvkit 540
QY 541 GHFYACQVARKTOEILTQVQHQQQKALQSGPPQSRKK 579
Db 541 ghfyacqvarktoeiltqvqhqqqkalksgppqsrkk 579

RESULT 2
AAB11365
ID AAB11365 standard; Protein; 579 AA.
XX
XX AAB11365;
XX
DT 21-FEB-2001 (first entry)

XX Human lung cancer associated antigen L523S.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
XX 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
XX Wang T, Fan L;
PI
DR WPI; 2000-628399/60.
DR N-PSDB; AAC66035.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -
XX
PS Claim 3; Page 259-261; 261pp; English.
XX
CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.
XX
SQ Sequence 579 AA;

Query Match 99.8%; Score 2943; DB 21; Length 579;
Best Local Similarity 99.7%; Pred. No. 6.5e-245;
Matches 577; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MNKLYIGNLSENAPDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKATEALS GK 60
Db 1 mnklyignlsenapsdlesifkdakipvsgpflvktgyafvdcpdsewalkatealsgk 60
QY 61 IELHGKPIEVHSEVPRQRIRKLIQIRNIPPHLQWEVLDSLLOVGVVESCSEQVNTDSETA 120
Db 61 ielhgkpievehsvprqrirklqirnipphlqwevldsllovgvvescseqvntdseta 120
QY 121 VVNVYSSKQDQARQALDKLNGFOLENTFLKVAYIPDDEMAAQNPLOQPRGRRLGQRGSS 180
Db 121 vvnvtysskqdarqaldklngfqlentflkvayipdemaagnploqprgrrlgqrgss 180
QY 181 RQSPGVSVKQKPCDPLRLVPTQFVGALIGEGATIRNITKTOSKIDVHRKENAGAA 240
Db 181 rgspgsvskqkpcdplrlvptqfvgaligegatirnitktoskldvhrkenagaa 240
QY 241 EKSTITLSTPEGTSAACKSILEIMHKEAQDIKFTTEEIPKLAHNNFVGRGLIGKEGRNLK 300
Db 241 eksitiltstpegtsaacksileimhkeaqdikftteeeipklahnnfvgrligrkegrnlk 300
QY 301 KIEQDQDTKITISPLQELTYLNPRTITVKGNVETCAKABEEIMKKIRESYENDIASMNL 360


```
Db 301 klegtdtkitispilqetllynpertitvkgvnetcakaeeimkkiresyendiamml 360
Qy 361 QAHLIPGLNALGLFPPTSGMPPPTSGPPSAMTPYPQEQSETEVHQRIPALSVGAI 420
Db 361 qahlipglinalglfpptsgmppsamtpypqeqsetetvhlfpalsvgai 420
Qy 421 IGKQGHKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPAQFKAQRIYKIKIENFV 480
Db 421 lkgqgghikqlsrfagasikiapaeapdakvrmvliitgppaqfkaqriykgikeenf 480
Qy 481 SPKEVKLEAHRVPSFAAGRVIGKGTVNELQNLSSAEVVPVPRDQPDENDQVVKIT 540
Db 481 spkeevkleahrpsfaagrvi gkgtvnelqnlssaevvprdqtpdendqv vkit 540
Qy 541 GHFYACQVAQRKIQEILTOVKHOQOKALQSGPPQSRRK 579
Db 541 ghfyacvaqrkiqeiltovkqhqqkqlsgppqsrirk 579

RESULT 3
ID AAY30649 standard; Protein; 577 AA.
XX
AC AAY30649;
XX
DT 17-NOV-1999 (first entry)
XX
DE A murine c-myc coding region determinant binding protein.
XX
KW c-myc coding region determinant binding protein; CRD-BP; tumor;
KW c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer.
XX
OS Mus musculus.
XX
PN WO9946594-A2.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US04897.
XX
PR 09-MAR-1998; 98US-0077372.
XX
PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI Ross J;
XX
DR WPI; 1999-551506/46.
XX
DR N-PSDB; AAZ10617.
XX
PT Diagnosing presence or absence of a tumor in a human by examining c-myc
XX coding region determinant-binding protein
XX
FS Example; Fig 1A-D; 79pp; English.
XX
CC The present sequence represents a murine c-myc coding region determinant
XX binding protein (CRD-BP). The presence or absence of a tumor can be
XX determined by determining the levels of CRD-BP present in the suspect
XX tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
XX and so prolongs its half-life. The methods are used for diagnosing
XX presence or absence of a tumor in a human, especially breast, colon
XX and pancreatic cancer. They are also used to inhibit cancer cell
XX growth.
XX
SQ Sequence 577 AA;

Query Match 74.1%; Score 2190; DB 20; Length 577;
Best Local Similarity 74.1%; Pred. No. 5.2e-180;
Matches 434; Conservative 74; Mismatches 16; Indels 8;

Qy 1 MNKLYIGNLSNAAPSDLESIFDKAIPVSGPFLVKTGAFVDCPDSEWALKATEALSGK 60
```

```
Db 1 mnllyignlsesvtpadlekfvfaehkisygqlvksyafvdcpdchamkaletfsgk 60
Qy 61 IELHGKPIEVEHSVKRQRIKLRNIPHLLOWEVLDSLLVQYGVVSECEQVNTDSETA 120
Db 61 velqgkrlemehsvpkkgrsklirnlppqlrwevldslaqgvtenceqvnteseta 120
Qy 121 VYNVYSSKQDARQALDKLNGFQLENFTLKVAYIPDEMAAQONPLQOPRRR-GLGQSGS 179
Db 121 vnyvysnregrtqaimkinghqlenhalksyipdeqitq---gpengrrtggfsg 176
Qy 180 SRQSGP---GSVSKQKPCDLRLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKEN 236
Db 177 prqgspsvaagapakqpvdiprlilvptqyvgailgkegatirnicktqskidvhrken 236
Qy 237 AGAAEKSITILPSTPETSAAKSIILEIMHKEAQDIKFTTEIPLKILAHNFFVRLIGKEG 296
Db 237 agaaekaisvhtpegcsackmilleimhkeakdtktadevpkilaahnfvrligk 296
Qy 297 RNLKKIEQDTRKITISPLQELTYNPERTITVKGNETCAKAEIEIMKKIRESYENDIA 356
Db 297 rnlkveqdtetkitisslqdlitlynpertitvkgaiencraeqeaimkkvreaeyenda 356
Qy 357 SMNLQAHLPGLNALGLFPPTSGMPPPTSGPPSAMT--PPYPQEQS-ETETVHQFIP 413
Db 357 amslqshlipglinalaavgilpasssavpp---ppssvtgaapyssfmqapeqemvqvfi 413
Qy 414 ALSVGAIIKQGOHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPAQFKAQRIYK 473
Db 414 aqvadailgkqghikqlsrfagasikiapetpdskvrmvviitgppaqfkaqriy 473
Qy 474 IKEENFVSPKEVKLEAHRVPSFAAGRVIGKGTVNELQNLSSAEVVPVPRDQPDEND 533
Db 474 lkeenfvpkeevkleahrpsfaagrvi gkgtvnelqnlssaevvprdqtpdend 533
Qy 534 QVVVKITGHFYACQVAQRKIQEILTOVKHOQOKALQSGPPQSRRK 579
Db 534 qvivkighfyasqmagrkirdilagvkq-qhqqkq-qsnaqarrk 577

RESULT 4
AAM38501
ID AAM38501 standard; Protein; 93 AA.
XX
AC AAM38501;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #12538 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, .Hauzel DK, Chen W, Rank DR;
XX
```

DR WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
PS Claim 27; SEQ ID No 38770; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 93 AA;
SQ
Query Match 16.0%; Score 474; DB 22; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.6e-33;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 135 ALOKINGFQLENFTLVKVAIYIPDEMAAQNPLQOPRRRLGQSGSRQSPGVSVKQKPC 194
Db 1 aldKngfqlenftlvkvaipdemaqqnplqprgrgrrglgrrgssrqqspgvsvkqkpc 60
QY 195 DLPLRLLVPTQFVGAIGKEGATIRNTKQTQS 227
Db 61 dlplrlvlptqfvgaiigkegatinrntkqtqs 93
RESULT 5
AAG03261
ID AAG03261 standard; Protein; 97 AA.
XX
AC AAG03261;
XX
DT 06-OCT-2000 (first entry)
XX
XX Human secreted protein, SEQ ID NO: 7342.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
KW
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST) GENSET.
PA
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR
XX N-PSDB; AAC03267.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT
XX
XX Claim 13; SEQ ID 7342; 71pp + CD-ROM; English.
PS
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CQ isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC

CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5',
CC and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 97 AA;
Query Match 11.5%; Score 341; DB 21; Length 97;
Best Local Similarity 64.6%; Pred. No. 7.9e-22;
Matches 62; Conservative 16; Mismatches 18; Indels 0; Gaps 0;
QY 1 MNKLYIGNLSENAAPSDELSIFKDAKIPVSGPFLVKTCYAFVDCPDSESWALKATEALSGK 60
Db 2 mnklyignlspavtaddrlqfgrdkipagqvlksgyafvdpdgqwaifraietlsgk 61
QY 61 IELHGKPIEVESHVSKRQIRKQLQIRNIPPHLOWEV 96
Db 62 vellhgkimevdysvskklsrskiqirnipphlqev 97
RESULT 6
AAR58813
ID AAR58813 standard; Protein; 644 AA.
XX
AC AAR58813;
XX
DT 13-APR-1995 (first entry)
XX
XX Human c-myc far upstream element (FUSE) binding protein (FBP).
DE
XX DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
KW promoter Pl.
KW
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 149
FT /label= M,I
XX
XX W09419465-A.
PN
XX
XX 01-SEP-1994.
PD
XX
XX 22-FEB-1994; 94WO-US01782.
PF
XX
XX 22-FEB-1993; 93US-0021608.
PR
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Avigan MI, Duncan RC, Levens DL;
PI
XX
XX WPI; 1994-294330/36.
DR
XX N-PSDB; AAQ68909.
XX
XX New DNA-binding regulator of c-myc expression and its cDNA - used
PT to develop prods. for diagnosis and therapy of disease states
PT such as tumour formation
PT
XX
XX Claim 5; Page 40-43; 94pp; English.
PS
XX DROME (DNA-binding regulator of c-myc expression) and FUSE (far
CC upstream element) binding protein (FBP) are synonymous. A FUSE which
CC is required for maximal transcription of c-myc binds a factor (DROME
CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
CC human c-myc promoter Pl. A full length FBP cDNA sequence was
CC assembled from overlapping clones obtd. from cDNA libraries. Source
CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
CC BJAB, and PMAP/PHA stimulated pooled human peripheral blood
CC lymphocytes (PBLS). A composite cDNA is given in AAQ68909 and the
CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in AAR58813.
CC

| XX | Sequence | 644 AA; |
|----|--|--|
| XX | Query Match | 8.2%; Score 241; DB 15; Length 644; |
| XX | Best Local Similarity | 22.3%; Pred. No. 6.1e-12; |
| XX | Matches 100; Conservative | 75; Mismatches 154; Indels 120; Gaps 17; |
| QY | 161 QQNPLQ---QPRRRGLGQRSGSGSPGSKOKPCDLPRLRLVPTQFVGAIICKEGAT | 217 |
| DB | 161 QQNPLQ---QPRRRGLGQRSGSGSPGSKOKPCDLPRLRLVPTQFVGAIICKEGAT | 217 |
| DB | 63 qkrplegdgdpdakkvapqndsfqtqlppmhqqgsrvmtteeykvpdgmvgfiigrgeq | 122 |
| QY | 218 IRNITKQTQKIDVHRKENAGAEEKSTITLSTPGTSAACKSILEIMHKEAODTKFTB-- | 275 |
| DB | 218 IRNITKQTQKIDVHRKENAGAEEKSTITLSTPGTSAACKSILEIMHKEAODTKFTB-- | 275 |
| DB | 123 isriqgesgckiqi-apsdsglperscxiltgtpesvqsakrllldqivekgrpapgfhhgd | 181 |
| QY | 276 ---EIPKILAHNNFVRLGKKEGRNLKKIBODTDTKITISPLQELTLYNPERT---- | 327 |
| DB | 276 ---EIPKILAHNNFVRLGKKEGRNLKKIBODTDTKITISPLQELTLYNPERT---- | 327 |
| DB | 182 gpnnavgeimipaskaglvlgkggetikqlqeragvkmvm--lqd---gpqntcgadkpl | 235 |
| QY | 328 TVKGNVETCAKAEIEIMKKIRE-----SYENDIASMNLQAHLPGLNLNALGLPPTSGM | 382 |
| DB | 328 TVKGNVETCAKAEIEIMKKIRE-----SYENDIASMNLQAHLPGLNLNALGLPPTSGM | 382 |
| DB | 236 ritgdpykvqakemvliellrdqdgffevrneygsr-----lqg-----negi | 278 |
| QY | 383 PPTSGPPSMTYPPOFEQSEETTVHQFTPALSVGAIIGQGHIKOLSRFAGASTKIA | 442 |
| DB | 383 PPTSGPPSMTYPPOFEQSEETTVHQFTPALSVGAIIGQGHIKOLSRFAGASTKIA | 442 |
| DB | 279 dvp-----iprfavgivigrngemikkikqndagvriqfk | 312 |
| QY | 443 PAEAPDAKVRVLIITGPE-AQFKAQ-----GRIYCK----- | 473 |
| DB | 443 PAEAPDAKVRVLIITGPE-AQFKAQ-----GRIYCK----- | 473 |
| DB | 313 pddgttpe-rlaqltgpddrcqhaaelitdlrrsvqagnpgpgpgrgrgrggnwnmg | 371 |
| QY | 474 ----IKENFVSPKKEVKELEAHIRVPSFAAGRVIGKGTQVNELQNLSAEVWVPRDQTP | 529 |
| DB | 474 ----IKENFVSPKKEVKELEAHIRVPSFAAGRVIGKGTQVNELQNLSAEVWVPRDQTP | 529 |
| DB | 372 ppgglqefnfi-----vpgktglgkgetiksisqsgsarieiqrnpqp | 418 |
| QY | 530 DENDQV-VVKITGH-----FYACQVAQRKI | 553 |
| DB | 530 DENDQV-VVKITGH-----FYACQVAQRKI | 553 |
| DB | 419 nadpnmkflftgtpqldyarlleeki | 447 |
| DB | 419 nadpnmkflftgtpqldyarlleeki | 447 |
| XX | RESULT 7 | |
| XX | AAG08900 | |
| ID | AAG08900 standard; Protein; 313 AA. | |
| XX | AC | |
| XX | AAG08900; | |
| XX | 17-OCT-2000 (first entry) | |
| XX | Arabidopsis thaliana protein fragment SEQ ID NO: 6617. | |
| DE | Protein identification; signal transduction pathway; metabolic pathway; | |
| KW | hybridisation assay; genetic mapping; gene expression control; promoter; | |
| KW | termination sequence. | |
| XX | Arabidopsis thaliana. | |
| OS | EP1033405-A2. | |
| PN | 06-SEP-2000. | |
| XX | 25-FEB-2000; 2000EP-0301439. | |
| XX | 25-FEB-1999; 99US-0121825. | |
| PR | 03-MAR-1999; 99US-0123380. | |
| PR | 09-MAR-1999; 99US-0123348. | |
| PR | 23-MAR-1999; 99US-0125788. | |
| PR | 25-MAR-1999; 99US-0126264. | |
| PR | 29-MAR-1999; 99US-0126785. | |
| PR | 01-APR-1999; 99US-0127462. | |
| PR | 06-APR-1999; 99US-0128234. | |
| PR | 08-APR-1999; 99US-0128714. | |
| PR | 16-APR-1999; 99US-0129845. | |
| PR | 19-APR-1999; 99US-0130077. | |
| PR | 21-APR-1999; 99US-0130449. | |
| PR | 23-APR-1999; 99US-0130510. | |
| PR | 23-APR-1999; 99US-0130899. | |
| PR | 28-APR-1999; 99US-0131449. | |
| PR | 30-APR-1999; 99US-0132048. | |
| PR | 30-APR-1999; 99US-0132407. | |
| PR | 04-MAY-1999; 99US-0132484. | |
| PR | 05-MAY-1999; 99US-0132485. | |
| PR | 06-MAY-1999; 99US-0132486. | |
| PR | 06-MAY-1999; 99US-0132487. | |
| PR | 07-MAY-1999; 99US-0132863. | |
| PR | 11-MAY-1999; 99US-0134256. | |
| PR | 14-MAY-1999; 99US-0134218. | |
| PR | 14-MAY-1999; 99US-0134219. | |
| PR | 14-MAY-1999; 99US-0134221. | |
| PR | 14-MAY-1999; 99US-0134370. | |
| PR | 18-MAY-1999; 99US-0134768. | |
| PR | 19-MAY-1999; 99US-0134941. | |
| PR | 20-MAY-1999; 99US-0135124. | |
| PR | 21-MAY-1999; 99US-0135353. | |
| PR | 24-MAY-1999; 99US-0135629. | |
| PR | 25-MAY-1999; 99US-0136021. | |
| PR | 27-MAY-1999; 99US-0136392. | |
| PR | 28-MAY-1999; 99US-0136782. | |
| PR | 01-JUN-1999; 99US-0137222. | |
| PR | 03-JUN-1999; 99US-0137528. | |
| PR | 04-JUN-1999; 99US-0137502. | |
| PR | 07-JUN-1999; 99US-0137724. | |
| PR | 08-JUN-1999; 99US-0138094. | |
| PR | 10-JUN-1999; 99US-0138540. | |
| PR | 10-JUN-1999; 99US-0138847. | |
| PR | 14-JUN-1999; 99US-0139119. | |
| PR | 16-JUN-1999; 99US-0139452. | |
| PR | 16-JUN-1999; 99US-0139453. | |
| PR | 17-JUN-1999; 99US-0139492. | |
| PR | 18-JUN-1999; 99US-0139454. | |
| PR | 18-JUN-1999; 99US-0139455. | |
| PR | 18-JUN-1999; 99US-013945 | |

| | | |
|----|--------------|---------------|
| PR | 19-JUL-1999; | 99US-0144335; |
| PR | 20-JUL-1999; | 99US-0144332; |
| PR | 20-JUL-1999; | 99US-0144632; |
| PR | 20-JUL-1999; | 99US-0144884; |
| PR | 21-JUL-1999; | 99US-0144814; |
| PR | 21-JUL-1999; | 99US-0145086; |
| PR | 21-JUL-1999; | 99US-0145088; |
| PR | 22-JUL-1999; | 99US-0145085; |
| PR | 22-JUL-1999; | 99US-0145087; |
| PR | 22-JUL-1999; | 99US-0145089; |
| PR | 22-JUL-1999; | 99US-0145192; |
| PR | 23-JUL-1999; | 99US-0145218; |
| PR | 23-JUL-1999; | 99US-0145224; |
| PR | 26-JUL-1999; | 99US-0145276; |
| PR | 27-JUL-1999; | 99US-0145913; |
| PR | 27-JUL-1999; | 99US-0145918; |
| PR | 28-JUL-1999; | 99US-0145919; |
| PR | 28-JUL-1999; | 99US-0145951; |
| PR | 02-AUG-1999; | 99US-0146386; |
| PR | 02-AUG-1999; | 99US-0146389; |
| PR | 02-AUG-1999; | 99US-0146388; |
| PR | 03-AUG-1999; | 99US-0147038; |
| PR | 04-AUG-1999; | 99US-0147204; |
| PR | 04-AUG-1999; | 99US-0147302; |
| PR | 05-AUG-1999; | 99US-0147192; |
| PR | 05-AUG-1999; | 99US-0147260; |
| PR | 06-AUG-1999; | 99US-0147303; |
| PR | 06-AUG-1999; | 99US-0147416; |
| PR | 09-AUG-1999; | 99US-0147493; |
| PR | 09-AUG-1999; | 99US-0147935; |
| PR | 10-AUG-1999; | 99US-0148171; |
| PR | 10-AUG-1999; | 99US-0148171; |
| PR | 11-AUG-1999; | 99US-0148319; |
| PR | 12-AUG-1999; | 99US-0148341; |
| PR | 13-AUG-1999; | 99US-0148565; |
| PR | 13-AUG-1999; | 99US-0148684; |
| PR | 16-AUG-1999; | 99US-0149368; |
| PR | 17-AUG-1999; | 99US-0149175; |
| PR | 18-AUG-1999; | 99US-0149426; |
| PR | 20-AUG-1999; | 99US-0149722; |
| PR | 20-AUG-1999; | 99US-0149723; |
| PR | 20-AUG-1999; | 99US-0149829; |
| PR | 23-AUG-1999; | 99US-0149902; |
| PR | 23-AUG-1999; | 99US-0149930; |
| PR | 25-AUG-1999; | 99US-0150566; |
| PR | 26-AUG-1999; | 99US-0150884; |
| PR | 27-AUG-1999; | 99US-0151065; |
| PR | 27-AUG-1999; | 99US-0151066; |
| PR | 27-AUG-1999; | 99US-0151080; |
| PR | 30-AUG-1999; | 99US-0151303; |
| PR | 31-AUG-1999; | 99US-0151438; |
| PR | 01-SEP-1999; | 99US-0151930; |
| PR | 07-SEP-1999; | 99US-0152363; |
| PR | 10-SEP-1999; | 99US-0153070; |
| PR | 13-SEP-1999; | 99US-0153758; |
| PR | 13-SEP-1999; | 99US-0154018; |
| PR | 16-SEP-1999; | 99US-0154399; |
| PR | 20-SEP-1999; | 99US-0154779; |
| PR | 23-SEP-1999; | 99US-0155139; |
| PR | 23-SEP-1999; | 99US-0155486; |
| PR | 24-SEP-1999; | 99US-0155659; |
| PR | 28-SEP-1999; | 99US-0156458; |
| PR | 29-SEP-1999; | 99US-0156596; |
| PR | 04-OCT-1999; | 99US-0157117; |
| PR | 05-OCT-1999; | 99US-0157765; |
| PR | 06-OCT-1999; | 99US-0157865; |
| PR | 07-OCT-1999; | 99US-0158029; |
| PR | 08-OCT-1999; | 99US-0158232; |
| PR | 12-OCT-1999; | 99US-0158369; |
| PR | 13-OCT-1999; | 99US-0159293; |
| PR | 13-OCT-1999; | 99US-0159294; |
| PR | 13-OCT-1999; | 99US-0159295; |
| PR | 14-OCT-1999; | 99US-0159329; |

| | | |
|----|--------------|---------------|
| PR | 14-OCT-1999; | 9905-0159330; |
| PR | 14-OCT-1999; | 9905-0159331. |
| PR | 14-OCT-1999; | 9905-0159637. |
| PR | 14-OCT-1999; | 9905-0159638. |
| PR | 14-OCT-1999; | 9905-0159639. |
| PR | 18-OCT-1999; | 9905-0159584. |
| PR | 21-OCT-1999; | 9905-0160784. |
| PR | 21-OCT-1999; | 9905-0160767. |
| PR | 21-OCT-1999; | 9905-0160768. |
| PR | 21-OCT-1999; | 9905-0160770. |
| PR | 21-OCT-1999; | 9905-0160814. |
| PR | 21-OCT-1999; | 9905-0160815. |
| PR | 22-OCT-1999; | 9905-0160816. |
| PR | 22-OCT-1999; | 9905-0160817. |
| PR | 22-OCT-1999; | 9905-0160818. |
| PR | 22-OCT-1999; | 9905-0160819. |
| PR | 22-OCT-1999; | 9905-0160820. |
| PR | 25-OCT-1999; | 9905-0161405. |
| PR | 25-OCT-1999; | 9905-0161406. |
| PR | 26-OCT-1999; | 9905-0161359. |
| PR | 26-OCT-1999; | 9905-0161360. |
| PR | 26-OCT-1999; | 9905-0161361. |
| PR | 28-OCT-1999; | 9905-0161920. |
| PR | 28-OCT-1999; | 9905-0161992. |
| PR | 28-OCT-1999; | 9905-0161993. |
| PR | 29-OCT-1999; | 9905-0162142. |

| | | | | |
|-----------------------|--------|--------------------|--------|-------------|
| Query Match | 7.8%; | Score 231.5; | DB 21; | Length 313; |
| Best Local Similarity | 25.1%; | Pred. No. 1.3e-11; | | |

| | | | |
|----|-----|---|-----|
| Qy | 183 | GSPGSVSKQKP-----CDLP--LRLLVPTQFVGALIGKAGATIRNITQTOSKID | 230 |
| | | : : : : : : : : : : : : : : : : | |
| Db | 10 | gspeelarkspethdsaeasxpthirflvnaaagsvigkggtitefqaksgariq | 69 |
| Qy | 231 | VHRKEN--AGAAEKSIITLSTPECTSAACKSIILEIMHKEAQDITKFTPEIP--LKILAHN | 285 |
| | | : : : : : : : : : : : : : : : : | |
| Db | 70 | lsrqneffpgttdrlimsgsikevungleilidklhsehaednevepriririvvpn | 129 |
| Qy | 286 | NFVGRLLTGKERNLKKTEQDTRKIIITSPLOELTLN--PERTITVGVNVTCAKEEIM | 344 |
| | | : : : : : : : : : : : : : : : : | |
| Db | 130 | sscgilgkggatkfsieeskagikspidn--tfvglsdrvltsgtfeeqmraidlil | 188 |
| Qy | 345 | KKIRESYENDIASNNLQAHILGILNALGLFPTSGMPPTSGPPSMTPTPPYPQEQSE | 404 |
| | | : : : : : : : : : : : : : : : | |
| Db | 189 | aklt---eddhysqnvhspsyaagynsvnyapngsg-----gkyqnkheea | 232 |
| Qy | 405 | TETVHQITPALSVCALIGKOGOHIKOLSRFAGASIKTAP-----AEPDAKVRMVIITGP | 459 |
| | | : : : : : : : : : : : : : : : | |
| Db | 233 | stvtivgaadehgivlgvrggrnimegtmarikisdrgdfmsgttdrkkvs---itgp | 289 |
| Qy | 460 | PEAQFKAAGRIYGKI | 474 |
| | | : : : : : : : : : : : : : : : | |
| Db | 290 | qraiqgaetmikkv | 304 |

| | | |
|----------|------------------|--|
| RESULT | 8 | |
| AAR58816 | | |
| ID | AAR58816 | standard; Protein; 590 AA. |
| XX | XX | |
| XX | AAR58816; | |
| XX | XX | |
| XX | XX | |
| XX | 13-APR-1995 | (first entry) |
| XX | XX | |
| DE | Human c-myc | far upstream element (FUSE), binding protein (FBP) |
| DE | variant from PBL | clone 31-10. |
| XX | XX | |
| KW | DNA-binding | regulator; c-myc; DRONE; FUSE; activator; cis-element; |
| KW | promoter P1. | |
| XX | XX | |
| OS | Homo sapiens. | |
| XX | XX | |
| PN | W09419465-A. | |
| XX | XX | |
| PD | 01-SEP-1994. | |

```

XX PF 22-FEB-1994; 94WO-US01782.
XX PR 22-FEB-1993; 93US-0021608.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Avigan MI, Duncan RC, Levens DL;
XX XX WPI; 1994-294330/36.
XX DR N-PSDB; AAQ68912.
XX XX
XX PT New DNA-binding regulator of c-myc expression and its cDNA - used
XX PT to develop prods. for diagnosis and therapy of disease states
XX PT such as tumour formation
XX PS Claim 20; Page 53-56; 94pp; English.
XX CC
XX CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
XX CC upstream element) binding protein (FBP) are synonymous. A FUSE which
XX CC is required for maximal transcription of c-myc binds a factor (DROME
XX CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
XX CC human c-myc promoter pl. A full length FBP cDNA sequence was
XX CC assembled from overlapping clones obtd.from cDNA libraries. Source
XX CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
XX CC BJAB, and PMA/PHA stimulated pooled human peripheral blood
XX CC lymphocytes (PBLs). A composite cDNA is given in AAQ68909 and the
XX CC deduced polypeptide sequence. (67.5 kd mol. wt.) is given in AAR58813.
XX CC The clone from an activated PBL cDNA library labeled "31-10" contains
XX CC an FBP variant. The ORF comprised of bps 135-1991 of AAQ68909 with
XX CC two exceptions. Clone 31-10 contains 63 bps inserted at posn. 238
XX CC which probably result from an intron which had not been spliced
XX CC out. The inserted bps remain in the ORF. The 31-10 clone also
XX CC deviates in that bps 1807-1952 are deleted. This deletion shifts
XX CC out of frame the stop codons which would terminate translation in
XX CC the other clones. The AA sequence for the protein encoded by clone
XX CC 31-10 can be found in AAR58816.
XX XX
XX SQ Sequence 590 AA;

```

```

Query Match 7.8%; Score 229.5; DB 15; Length 590;
Best Local Similarity 22.5%; Pred. No. 5.2e-11;
Matches 91; Conservative 69; Mismatches 128; Indels 117; Gaps 16;

```

```

Qy 202 VPTQFVGALCKEGATIRNITKQTSKIDVHRKENAGAAEKSIITLSPEGTSACKSIL 261
Db 91 vpdgmvgfiirggedisrqqesgckiqi- apdsggiperscmiltgtpesvqsakrild 149
Qy 262 RIMHKEAQDIKFTE-----EIPKTLAHNFVGRIGKEGRNLKKEIDTDTKITISPLQ 316
Db 150 qivekgrpapfhgdpgpnnavqeimipaskagivigkggtikqlqragvkmvm--iq 207
Qy 317 ELTYNPRT-----ITVKGNVETCAKAEIRMKKIRE-----SYENDIASMNLQAHLP 366
Db 208 d-----gpqntgadkplritgdykvqgakenvlelirdqgqfrevrneygsr-----ig 257
Qy 367 GLNLNALGLFPTSCMPPTSGPSAMTPPYQPEQSETEVHFQIPALSVGAIIKGQ 426
Db 258 g-----negidvp-----iprfavgivigrnge 280
Qy 427 HIKOLSRFAGASIKIAPAEAPDAKVRMVIITGPPPE-AQFKAQ----- 467
Db 281 mikkiqndagvriqkpdgttpe-riagitgppdrcdhaeaitdlrrsvqagnpqpg 339
Qy 468 --GRIYK-----IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTVNEL 513
Db 340 pggirgrgqgnwmppggigqefnfi-----vptgtglligkgetiksi 386
Qy 514 QNLSSAEVWVRDQTPDENDOV-VVKITGH-----FYACQVAQRKI 553
Db 387 sqqsqgarielrnpnpnadpnmkiftirgtppqidiyarqlleeki 431

```

```

RESULT 9
AAR58814
ID AAR58814 standard; Protein; 643 AA.
XX AC AAR58814;
XX DT 13-APR-1995 (first entry)
XX DE Human c-myc far upstream element (FUSE) binding protein (FBP)
XX DE variant lacking Ser 97.
XX KW DNA-binding regulator; c-myc; DROME; FUSE; activator; cis-element;
XX KW promoter pl.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 148
XX FT /label= M,I
XX PN W09419465-A.
XX PD 01-SEP-1994.
XX PF 22-FEB-1994; 94WO-US01782.
XX PR 22-FEB-1993; 93US-0021608.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Avigan MI, Duncan RC, Levens DL;
XX DR WPI; 1994-294330/36.
XX DR N-PSDB; AAQ68910.
XX PT New DNA-binding regulator of c-myc expression and its cDNA - used
XX PT to develop prods. for diagnosis and therapy of disease states
XX PT such as tumour formation
XX PS Claim 12; Page 60-63; 94pp; English.
XX CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
XX CC upstream element) binding protein (FBP) are synonymous. A FUSE which
XX CC is required for maximal transcription of c-myc binds a factor (DROME
XX CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
XX CC human c-myc promoter pl. A full length FBP cDNA sequence was
XX CC assembled from overlapping clones obtd.from cDNA libraries. Source
XX CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
XX CC BJAB, and PMA/PHA stimulated pooled human peripheral blood
XX CC lymphocytes (PBLs). A composite cDNA is given in AAQ68909 and the
XX CC deduced polypeptide sequence. (67.5 kd mol. wt.) is given in AAR58813.
XX CC Three clones from a BJAB cDNA library and three clones from an
XX CC activated human PBL cDNA library contain the sequence in AAQ68909.
XX CC Three clones from the same BJAB library and two clones from the
XX CC activated lymphocyte library are lacking bps 316,317 and 318
XX CC (see AAQ68910). The mRNA lacking these three bps would encode a
XX CC variant protein lacking Ser 97 (see AAR58814).
XX SQ Sequence 643 AA;

```

```

Query Match 7.7%; Score 228.5; DB 15; Length 643;
Best Local Similarity 22.0%; Pred. No. 7.3e-11;
Matches 99; Conservative 77; Mismatches 152; Indels 121; Gaps 18;

```

```

Qy 161 QONPLQ---QPRRRGLGQRSGSGSVSKQKPCDPLRLVPTQFVGAIIGEGAT 217
Db 63 qkrpiedgqpdakvqpnqdsfgtqlp-pmhqgqrsvmteeykvpdgmvgfiirgqeq 121
Qy 218 IRNITKQTSKIDVHRKENAGAAEKSIITLSPEGTSACKSILEIMHKEADIKFTE-- 275
Db 122 lsriqgesgckiqi-apdsggiperscxitgtpesvqsakrildgivekgrpagfhgd 180

```


XX AAB42547;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF2311 polypeptide sequence SEQ ID NO:4622.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antineumatic; antithyroid;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; erythematosis; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shinkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC76756.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 3823-3824; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antiparasitic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antineumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 403 AA;

Query Match 7.2%; Score 213.5; DB 21; Length 403;
Best Local Similarity 25.6%; Pred. No. 6.9e-10;
Matches 81; Conservative 61; Mismatches 126; Indels 49; Gaps 13;
QY 183 GSPGSVSKQPCD--LPLRLVPTQFVGAIIKGEATIRNITKOTSKIDVHRKENAGAA 240
DB 3 gsdggleeepelsitlrlmlhmgkevsllgkgetvkrirgqsariti---segsc 58
QY 241 EKSITLSPGTSAAKSKSILEIMHKEAQDI-----KETEELPLKTLAHNNFVGR 291
DB 59 perittit--gstaavfhavsmiafkldeicaapangnvsrppvrlrivi pascgsl 116
QY 292 IGKEGNLKKIEQDQDTKITISPLQELTYNPRTITVKNVETCAKAEIEIMKKIRESY 351
DB 117 igkagtkikeirettgaqgva--gdllpnsteravtvsqvpdaillcvrqicavilles 174
QY 352 ENDIASMNIQALHLPGLNLNALGLFPPTSGMPPTSGPPSAMTP-----PY 397
DB 175 pk---gatipyh--pslsigtv--lls andgf--svdggvgavtpaevtklqq lsshavpf 226
QY 398 -----PQFEQ-SETETVHQFIPALSVGALIGKQGHQIKLSRFAGASIKIAPAEAPDAK 450
DB 227 atpsvvpvgldpqtqtssqeflvndligcvigrqskiseirgmsghhikig-nqaeag 285
QY 451 VRMVIITGPPEAQFRAQ 467
DB 286 erhvtitgspvsialaq 302
RESULT 14
AAB94825
ID AAB94825 standard; Protein; 403 AA.
XX AAB94825;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:15981.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 15981; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of

| | | | |
|-----------------|--|---|---------------|
| PR 13-JUL-1999; | 99US-0143542. | PR 04-OCT-1999; | 99US-0157117. |
| PR 14-JUL-1999; | 99US-0143624. | PR 05-OCT-1999; | 99US-0157753. |
| PR 15-JUL-1999; | 99US-0144005. | PR 06-OCT-1999; | 99US-0157865. |
| PR 16-JUL-1999; | 99US-0144085. | PR 07-OCT-1999; | 99US-0158029. |
| PR 16-JUL-1999; | 99US-0144086. | PR 08-OCT-1999; | 99US-0158232. |
| PR 19-JUL-1999; | 99US-0144325. | PR 12-OCT-1999; | 99US-0158369. |
| PR 19-JUL-1999; | 99US-0144331. | PR 13-OCT-1999; | 99US-0159293. |
| PR 19-JUL-1999; | 99US-0144332. | PR 13-OCT-1999; | 99US-0159294. |
| PR 19-JUL-1999; | 99US-0144333. | PR 13-OCT-1999; | 99US-0159295. |
| PR 19-JUL-1999; | 99US-0144334. | PR 14-OCT-1999; | 99US-0159329. |
| PR 20-JUL-1999; | 99US-0144335. | PR 14-OCT-1999; | 99US-0159330. |
| PR 20-JUL-1999; | 99US-0144352. | PR 14-OCT-1999; | 99US-0159331. |
| PR 20-JUL-1999; | 99US-0144632. | PR 14-OCT-1999; | 99US-0159637. |
| PR 20-JUL-1999; | 99US-0144884. | PR 14-OCT-1999; | 99US-0159638. |
| PR 21-JUL-1999; | 99US-0144814. | PR 18-OCT-1999; | 99US-0159584. |
| PR 21-JUL-1999; | 99US-0145086. | PR 21-OCT-1999; | 99US-0160741. |
| PR 21-JUL-1999; | 99US-0145088. | PR 21-OCT-1999; | 99US-0160767. |
| PR 22-JUL-1999; | 99US-0145085. | PR 21-OCT-1999; | 99US-0160768. |
| PR 22-JUL-1999; | 99US-0145087. | PR 21-OCT-1999; | 99US-0160770. |
| PR 22-JUL-1999; | 99US-0145089. | PR 21-OCT-1999; | 99US-0160814. |
| PR 22-JUL-1999; | 99US-0145192. | PR 21-OCT-1999; | 99US-0160815. |
| PR 23-JUL-1999; | 99US-0145145. | PR 22-OCT-1999; | 99US-0160980. |
| PR 23-JUL-1999; | 99US-0145218. | PR 22-OCT-1999; | 99US-0160981. |
| PR 23-JUL-1999; | 99US-0145224. | PR 22-OCT-1999; | 99US-0160989. |
| PR 26-JUL-1999; | 99US-0145276. | PR 25-OCT-1999; | 99US-0161404. |
| PR 27-JUL-1999; | 99US-0145913. | PR 25-OCT-1999; | 99US-0161405. |
| PR 27-JUL-1999; | 99US-0145918. | PR 25-OCT-1999; | 99US-0161406. |
| PR 27-JUL-1999; | 99US-0145919. | PR 26-OCT-1999; | 99US-0161359. |
| PR 28-JUL-1999; | 99US-0145951. | PR 26-OCT-1999; | 99US-0161360. |
| PR 02-AUG-1999; | 99US-0146386. | PR 26-OCT-1999; | 99US-0161361. |
| PR 02-AUG-1999; | 99US-0146388. | PR 28-OCT-1999; | 99US-0161920. |
| PR 02-AUG-1999; | 99US-0146389. | PR 28-OCT-1999; | 99US-0161992. |
| PR 03-AUG-1999; | 99US-0147038. | PR 28-OCT-1999; | 99US-0161993. |
| PR 04-AUG-1999; | 99US-0147204. | PR 29-OCT-1999; | 99US-0161993. |
| PR 04-AUG-1999; | 99US-0147302. | PR 29-OCT-1999; | 99US-0162142. |
| PR 05-AUG-1999; | 99US-0147192. | | |
| PR 05-AUG-1999; | 99US-0147260. | | |
| PR 06-AUG-1999; | 99US-0147303. | | |
| PR 06-AUG-1999; | 99US-0147416. | | |
| PR 09-AUG-1999; | 99US-0147493. | | |
| PR 09-AUG-1999; | 99US-0147935. | | |
| PR 10-AUG-1999; | 99US-0148171. | | |
| PR 11-AUG-1999; | 99US-0148319. | | |
| PR 12-AUG-1999; | 99US-0148341. | | |
| PR 13-AUG-1999; | 99US-0148565. | | |
| PR 13-AUG-1999; | 99US-0148684. | | |
| PR 16-AUG-1999; | 99US-0149368. | | |
| PR 17-AUG-1999; | 99US-0149175. | | |
| PR 18-AUG-1999; | 99US-0149426. | | |
| PR 20-AUG-1999; | 99US-0149722. | | |
| PR 20-AUG-1999; | 99US-0149723. | | |
| PR 20-AUG-1999; | 99US-0149929. | | |
| PR 23-AUG-1999; | 99US-0149902. | | |
| PR 23-AUG-1999; | 99US-0149930. | | |
| PR 25-AUG-1999; | 99US-0150566. | | |
| PR 26-AUG-1999; | 99US-0150884. | | |
| PR 27-AUG-1999; | 99US-0151065. | | |
| PR 27-AUG-1999; | 99US-0151066. | | |
| PR 27-AUG-1999; | 99US-0151080. | | |
| PR 30-AUG-1999; | 99US-0151303. | | |
| PR 31-AUG-1999; | 99US-0151438. | | |
| PR 01-SEP-1999; | 99US-0151930. | | |
| PR 07-SEP-1999; | 99US-0152363. | | |
| PR 13-SEP-1999; | 99US-0153070. | | |
| PR 13-SEP-1999; | 99US-0153758. | | |
| PR 15-SEP-1999; | 99US-0154018. | | |
| PR 16-SEP-1999; | 99US-0154039. | | |
| PR 20-SEP-1999; | 99US-0154779. | | |
| PR 22-SEP-1999; | 99US-0155139. | | |
| PR 23-SEP-1999; | 99US-0155486. | | |
| PR 24-SEP-1999; | 99US-0155659. | | |
| PR 28-SEP-1999; | 99US-0156458. | | |
| PR 29-SEP-1999; | 99US-0156596. | | |
| | | Query Match 6.9%; Score 204.5; DB 21; Length 781; | |
| | | Best Local Similarity 19.4%; Pred. No. 1.le-08; | |
| | | Matches 93; Conservative 82; Mismatches 177; Indels 127; Gaps 15; | |
| QY 199 | RLVPTQFVGAIGKEGATIRNITKOTSKIDVHRKENAGAAEKSTITLSTPEGTSAAK 258 | | |
| Db 282 | rlcpidvvgvgigsgkvinairhtkakikvfdqlh-gcsqrvtiycs----- 331 | | |
| QY 259 | SILEIMHKEAQDIKFT--BEIPL-----KI 281 | | |
| Db 332 | -----vkeqeelgftksenepliccaqdaallkvydaivasdeenctknvdrdnkcerl 386 | | |
| QY 282 | LAHNFVGRLLIGKEGRNLKIEQDUTKI-----TISPIQELTYLPERTITVKGNVETC 336 | | |
| Db 387 | lvpfsgssllgkagenikrirttrrasvkvsdkvdsphvcameydnvvvisgepesv 446 | | |
| QY 337 | AK---AEEETMKKI--RESYENDIASMNL-----QAHL--- 364 | | |
| Db 447 | kqalfavsaimykinprenipldstsqdvpaaasviypsdlsnsvypqtgfyysnqdhllq 506 | | |
| QY 365 | ---IPGLNINLGL-----FPPTSGMPPPTSGPPSAMTPPYQPQEOSETETVHQFIPALS 416 | | |
| Db 507 | gaqvpsy-fnalsvsdfgyaetaanpvvfasslpythgfggssrseelvfkvlclpcn 565 | | |
| QY 417 | VGAIGKQGOHIKQLSRFACASIKIAPAEAPDAKVRWVIITGPPEAQF----- 464 | | |
| Db 566 | imrvlgkggstikrreagsgcielevndsrctkcgddceviivtateillfcclstpfvfmg 625 | | |
| QY 465 | ----KAQGRYKIKEENFVSPKEEVKLEAHIRVPFAAGRVIGKGTVNEQLNLSAE 520 | | |
| Db 626 | pddmksma-veavllliqeyindedaenvkmqlilvsskvgicvkgsgsvineirkrnan 684 | | |
| QY 521 | VVVPDRDQTPDENDQVVVKITGHFYACQVAKRQKIQEILLTOVKHQKQKALOSGPPQSRK 579 | | |
| Db 685 | iciskgkdd-----lvevsgev-----ssvrdaliqivlrredvlgdkdsvatr 731 | | |

Search completed: January 25, 2002, 22:11:55
Job time: 26176 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 15:00:14 ; Search time 247.67 Seconds
(without alignments)
52.608 Million cell updates/sec

Title: US-09-685-696-176

Perfect score: 2956

Sequence: 1 MNKLYIGNLSENAPSDLES.....VKHQOQKALQSGPPQSRRK 579

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-------------------|--------------------|
| 1 | 2190 | 74.1 | 577 | 4 | US-09-261-855-2 | Sequence 2, Appli |
| 2 | 241 | 8.2 | 644 | 1 | US-08-021-608D-2 | Sequence 2, Appli |
| 3 | 241 | 8.2 | 644 | 1 | US-08-726-160-2 | Sequence 2, Appli |
| 4 | 241 | 8.2 | 644 | 5 | PCT-US94-01782-2 | Sequence 2, Appli |
| 5 | 238 | 8.1 | 49 | 4 | US-09-261-855-22 | Sequence 22, Appli |
| 6 | 237 | 8.0 | 48 | 4 | US-09-261-855-24 | Sequence 24, Appli |
| 7 | 236.5 | 8.0 | 643 | 1 | US-08-021-608D-10 | Sequence 10, Appli |
| 8 | 236.5 | 8.0 | 643 | 1 | US-08-726-160-10 | Sequence 10, Appli |
| 9 | 236.5 | 8.0 | 643 | 5 | PCT-US94-01782-10 | Sequence 10, Appli |
| 10 | 232 | 7.8 | 49 | 4 | US-09-261-855-18 | Sequence 18, Appli |
| 11 | 229.5 | 7.8 | 590 | 1 | US-08-021-608D-8 | Sequence 8, Appli |
| 12 | 229.5 | 7.8 | 590 | 1 | US-08-726-160-8 | Sequence 8, Appli |
| 13 | 229.5 | 7.8 | 590 | 5 | PCT-US94-01782-8 | Sequence 8, Appli |
| 14 | 227 | 7.7 | 47 | 4 | US-09-261-855-23 | Sequence 23, Appli |
| 15 | 227 | 7.7 | 530 | 1 | US-08-187-793-4 | Sequence 4, Appli |
| 16 | 225 | 7.6 | 47 | 4 | US-09-261-855-21 | Sequence 21, Appli |
| 17 | 219 | 7.4 | 48 | 4 | US-09-261-855-20 | Sequence 20, Appli |
| 18 | 218 | 7.4 | 47 | 4 | US-09-261-855-17 | Sequence 17, Appli |
| 19 | 190 | 6.4 | 47 | 4 | US-09-261-855-19 | Sequence 19, Appli |
| 20 | 159.5 | 5.4 | 471 | 4 | US-08-866-928B-1 | Sequence 1, Appli |
| 21 | 154 | 5.2 | 243 | 1 | US-08-187-793-2 | Sequence 2, Appli |
| 22 | 147 | 5.0 | 243 | 1 | US-08-021-608D-6 | Sequence 6, Appli |
| 23 | 147 | 5.0 | 243 | 1 | US-08-726-160-6 | Sequence 6, Appli |
| 24 | 147 | 5.0 | 243 | 5 | PCT-US94-01782-6 | Sequence 6, Appli |
| 25 | 146.5 | 5.0 | 688 | 4 | US-08-973-273-26 | Sequence 26, Appli |
| 26 | 146.5 | 5.0 | 747 | 4 | US-08-973-273-3 | Sequence 3, Appli |
| 27 | 145.5 | 4.9 | 414 | 1 | US-07-667-276A-4 | Sequence 4, Appli |

ALIGNMENTS

RESULT 1

US-09-261-855-2

; Sequence 2, Application US/09261855A

; Patent No. 6255055

; GENERAL INFORMATION:

; APPLICANT: Ross, Jeffrey

; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN

; FILE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE

; FILE REFERENCE: 960296.95131

; CURRENT APPLICATION NUMBER: US/09/261,855A

; CURRENT FILING DATE: 1999-03-03

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 577

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-261-855-2

| | | | | |
|-----------------------|------------------|---|------------------------|-------------|
| Query Match | 74.1% | Score 2190; | DB 4; | Length 577; |
| Best Local Similarity | 74.1% | Pred. No. 1.4e-187; | | |
| Matches 434; | Conservative 62; | Mismatches 74; | Indels 16; | Gaps 8; |
| Qy | 1 | MNKLYIGNLSENAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDDES | WALKAEIALSGK | 60 |
| Db | 1 | MNKLYIGNLSESVTPADLEKVFPAEHKISYSGQPLVKS | GVAFVDCPDHWMAMKAETPSGK | 60 |
| Qy | 61 | IELHGKPIEVEHSVPKQRIKQIRNIPPHLOWEVLDSLLVQYGVVSECEQVNTDSETA | 120 | |
| Db | 61 | VELQGRLEMEHSVPKQRSRKIQIRNIPQLRWEVLDSLLAQYGVENEQVNTSETA | 120 | |
| Qy | 121 | VNVYSSKDOARQALDKLNGFOLENTLKVAYIPDEMAAQNPLOQPRGR | -GLGORG | 179 |
| Db | 121 | VNVYTSREQTRQALMKLNGHOLENHALKVSYIPDEQITQ | ----GPENRRGGFSRGQ | 176 |
| Qy | 180 | SROGSP---GSVSKQKPCDPLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKEN | 236 | |
| Db | 177 | PROGSPVAGAPAKQPVDPILRLVPTQYVGAIIGKEGATIRNITKQTSKIDVHRKEN | 236 | |
| Qy | 237 | AGAAEKSTILTPEGTSACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVRLIGKEG | 296 | |
| Db | 237 | AGAAEKAIKSVHSTPEGSSACKMILEIMHKEAKDTADEVPLKILAHNNFVRLIGKEG | 296 | |
| Qy | 297 | RNLKKEQDFTKITISPLQELTYLPERTITVKNVETCAKAEETIMKIRSYENDIA | 356 | |
| Db | 297 | RNLKKEQDFTKITISSLODLYLPERTITVKGAIENCCRAEQEIMKVRAYENDVA | 356 | |
| Qy | 357 | SMNLQAHLPGLNMLALGFLPPTSGMPPPTSGPPSANT--PPYQFEQS--ETETVHQFIP | 413 | |

| | | | |
|----|-----|--|-----|
| Db | 357 | AMSLQSHLIPGLNLAANGVIFPASPSSAVPP-PPSSVVTGAAVPSFQMGAPEQMVQVFIP | 413 |
| QY | 414 | ALSVGAIIGKGGHIKQLSRFAGASIKIAPAPADAKVRWITITGPPEAQFKAQRIYKG | 473 |
| Db | 414 | AOAVGAIIGKGGHIKQLSRFASASIKIAPPETPSKVRVMVITGPPEAQFKAQRIYKG | 473 |
| QY | 474 | IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVELQNLSSAEVYVVPDQTPDEND | 533 |
| Db | 474 | LKEENFFGPKKEEVKLETHIRVPASAGRVIGKGGKTVELQNLTAEEVYVVPDQTPDEND | 533 |
| QY | 534 | QVYVKITGHFYACQVAKRIQELLTOVKHQHQOKALQSGPPQSRKK | 579 |
| Db | 534 | QVTVKIIGHFYASOMAKRIRDTLQVKG-QHQKGG-QSNLAQARRK | 577 |

RESULT 2
US-08-021-608D-2
: Sequence 2, Application US/08021608D
: Patent No. 5580760
: GENERAL INFORMATION:
: APPLICANT: LEVENS, DAVID L., DUNCAN,
: APPLICANT: ROBERT C., AND AVIGAN, MARK I.
: TITLE OF INVENTION: NOVEL FUSE BINDING
: TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR

Query Match 8.2%; Score 241; DB 1; Length 644;
Best Local Similarity 22.3%; Pred. NO. 4.2e-13;
Matches 100; Conservative 75; Mismatches 154; Indels 1

| | | | |
|----|-----|---|-----|
| Qy | 161 | QONPLQ---OPRRRRGLQGRSSRQSGVSKOKPCDPLRLLLVPTQFVGAIIIGKEGAT | 217 |
| Db | 63 | KQRPLEDQDQDAKKVAPQNDSEFTQLPPMHQQQSRQSVMTTEYKVPDGMVFTIIGRGEQ | 122 |
| Qy | 218 | IRNITTKQTSKIDVHRKENAGAEEKSITILSTPEGTSAAKCSITLEIMHKEAQDIKFE- | 275 |
| Db | 123 | ISRIQQESGCKIQI-APDSGGLPERSCXJGTPESVQSAKRLDQIIVEKRPAPGPHHG | 181 |
| Qy | 276 | --EIPULKLIAHNNFVGRGLICEGRNLKKTIEQDTDTKITISPLOELTLXNPET---- | 327 |
| Db | 182 | PGNAVQEIIMPASKAGLVIGKGETIKQLQERAGVKVM--IQD---GPQNTGADKPL | 235 |
| Qy | 328 | TVKGNVETCAKABEEIMKKTRE-----SYENDIASNQLAHILPGLNALGLFPPTSGM | 382 |
| Db | 236 | RITGDPYKVQOAKEMWLELIRDOGGPREVNEYSGR-----ICG-----NEGI | 278 |
| Qy | 383 | PPPTSGPPSAMPYPYQFQESETVHQFIPALSVGAIGKQGOHIKQLSRFAGASIKIA | 442 |
| Db | 279 | DVP-----IPRAVGIVIGRNGEMIKKIQNDAGVRIQFK | 312 |
| Qy | 443 | PAEAPDAKVRWIIITGPPE-AQPKAQ-----GRYKG----- | 473 |
| Db | 313 | PDDGTTPE-RIAQITGPPDRCQHAAEIITDLLRSVOAGNPGPGPGRGRGQGNWNG | 371 |
| Qy | 474 | ----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGGGKTVNELQNLISAEVVVPRDQTP | 529 |
| Db | 372 | PPGGLQBFNI-----VPTGKTGLIIGKGETIKSISQSGARIELQRNPPP | 418 |
| Qy | 530 | DENDQV-VVKITGH---FYACQAVQARKI | 553 |
| Db | 419 | NADPNMKLFTIRGTPPOIDIVAROLIEEKI | 447 |

RESULT 3
US-08-726-160-2
; Sequence 2, Application US/08736160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 644


```

Db 371 PPGGLQEFNFI-----VPTGKTGLIIGKGETIKSIQSQSGARIELQRNPPP 417
Qy 530 DENDQV-VVKITGH-----FYACQVAQRKI 553
Db 418 NADPNMKLFTIRGTPOQIDYARQLIEKI 446

RESULT 8
US-08-726-160-10
; Sequence 10, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: NOVEL FUSE BINDING
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726.160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: Amino Acid 148 (Xaa) is Met or Ile
;
US-08-726-160-10
Query Match 8.0%; Score 236.5; DB 1; Length 643;
Best Local Similarity 22.3%; Pred No. 1.1e-12;
Matches 100; Conservative 77; Mismatches 151; Indels 121; Gaps 18;

Qy 161 QQNPLQ---QPRRRGLGQSSRGSGSVSKQKPCDLPRLRLVPTQFVGAIIGKEGAT 217
Db 63 QKRPLEDGPQDAKKVAPQNDSEFGTLP-PMHQQORSVMTVEYKVPDGMVGFIIIGRGQ 121
Qy 218 INRITKQTSKIDVRHKNAGAAEKSIITLSTPEGTSACKSILEIMHKEAQDKPTE-- 275
Db 122 ISRIQOESGCKIQI-APDSGGLPERSCXLTGTPESVQSAKRLLDQIVEKGRPAPGFHHGD 180
```

```

Qy 276 ---EIPKLIAHNHNVGRILIGKEGRNLKKIEQDTRKTIITISPLQELTLYNPRT-----I 327
Db 181 GPGNAVQEIWIPIASKAGLVIGKGETIKQLOERAGVKVMW--IQD---GPQNTGADKPL 234
Qy 328 TVKGNVETCAKAEIEIMKKIRE-----SYENDIASNNLOAHLPGLNGLNLFPPPTSGM 382
Db 235 RITGDPYKVVQAKEMVLELIRDOGGFREVNRVEYGSR-----IGG-----NEGI 277
Qy 383 PPPTSGPPSAMTPPYQPFQSETEVHQIFALSVGAIIICKQGHKQLKLSRFAGASIKIA 442
Db 278 DVP-----IPFAVGIVIGRNGEMIKKIQNDAGVRIQFK 311
Qy 443 PAEAPDAKVRMVIITGPPE-AQFKAO-----GRIYVK-----473
Db 312 PDGGITPE-RIAQITGPPDRCCQHAELIITDLRSVQAGNPGPGGGRGRGQGNWNG 370
Qy 474 ----TKEENFVSPKEVKLEAHIRVPSFAAGRVIGKGTVTNELQNLSSAEVVVVRDQTP 529
Db 371 PPGGLQEFNFI-----VPTGKTGLIIGKGETIKSIQSQSGARIELQRNPPP 417
Qy 530 DENDQV-VVKITGH-----FYACQVAQRKI 553
Db 418 NADPNMKLFTIRGTPOQIDYARQLIEKI 446

RESULT 9
PCT-US94-01782-10
; Sequence 10, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; STRAIN:
```


Db 208 D-----GFQNTGADRLRITGDPYKQAKEMVLEIRDOGGFREVRNEYSR-----IG 257
QY 367 GLNLNALGLFPPTSGMPPTSGPPSAMTPPYQPEQSETEVHOFIPALSGAIIKGQO 426
Db 258 G-----NEGIDVP-----IPREAVGIVIGRNGE 280
QY 427 HIKLSRFAGASIKIAPAEAPDAKVRMVIITGPPE-AQFKAQ----- 467
Db 281 MIKKIQNDAGVRIQFKPDGTTPE-RIAQITGPPDRCHAAEIIITDLRSVQAGNPGPG 339
QY 468 --GRIYK-----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTVNEL 513
Db 340 PGRGRGRGOGNWMNMGPPGLQEFNF-----VPTGKTGLIIGKGETIKSI 386
QY 514 QNLSSAEVVVPRDQTPDENOV-VVKITGH-----FYACQVAORKI 553
Db 387 SQSGARIELORPPPNADPNMKLFTINGTPOQIDYARQLIEEKI 431

RESULT 12

US-08-726-160-8
; Sequence 8, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
US-08-726-160-8

Query Match 7.8%; Score 229.5; DB 1; Length 590;
Best Local Similarity 22.5%; Pred. No. 3.9e-12;

Matches 91; Conservative 69; Mismatches 128; Indels 117; Gaps 16;
QY 202 VPTQFVGAIIGKEGATIRNITKQTSKIDVHRKENAGAEKSITILSPGTSAAACKSIL 261
Db 91 VPDGMVFIIGRGEQISRIOESCKIQI-APDSGGLPERSCLMTGTPEVSQSAKRLLD 149
QY 262 EIMHKEAQDIKFE-----EIPKLIAHNNEFVGRILIGKGRNLKKIEQDTDTKITISPLQ 316
Db 150 QIVEKGRPAPGFHGDGPGNAVQEIIMPASKAGLVIGKGETIKQLQERAGVKVMW--IQ 207
QY 317 ELTYLNPERT-----ITVKNVETCAKAEELMKKIRE-----SYENDIASNHLQALIP 366
Db 208 D-----GPQNTGADRLRITGDPYKQAKEMVLEIRDOGGFREVRNEYSR-----IG 257
QY 367 GLNLNALGLFPPTSGMPPTSGPPSAMTPPYQPEQSETEVHOFIPALSGAIIKGQO 426
Db 258 G-----NEGIDVP-----IPREAVGIVIGRNGE 280
QY 427 HIKLSRFAGASIKIAPAEAPDAKVRMVIITGPPE-AQFKAQ----- 467
Db 281 MIKKIQNDAGVRIQFKPDGTTPE-RIAQITGPPDRCHAAEIIITDLRSVQAGNPGPG 339
QY 468 --GRIYK-----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTVNEL 513
Db 340 PGRGRGRGOGNWMNMGPPGLQEFNF-----VPTGKTGLIIGKGETIKSI 386
QY 514 QNLSSAEVVVPRDQTPDENOV-VVKITGH-----FYACQVAORKI 553
Db 387 SQSGARIELORPPPNADPNMKLFTINGTPOQIDYARQLIEEKI 431

RESULT 13

PCT-US94-01782-8
; Sequence 8, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown

MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE: Human
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: HL60
ORGANELLE:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PCT-US94-01782-8

Query Match 7.8%; Score 229.5; DB 5; Length 590;
Best Local Similarity 22.5%; Pred. No. 3.9e-12;
Matches 91; Conservative 69; Mismatches 128; Indels 117; Gaps 16;

QY 202 VPTQFYCAIIGKCATIRNITKQTSKIDVHRKENAGAAEKSTITLSTPGETSAACKSIL 261
Db 91 VPDGMVGIIGRGGEOLSRISQESGCKIQI-APDSGGLPERSCLMTGTGPESVQSAKRLD 149

QY 262 EIMHKAQDIKFTF-----EIPKILAHNNFVGRGLIGKEGRNLKTEQDITDITISPLQ 316
Db 150 QIVEKGRPAPFHHGDCPGNAVEIMIPASKAGLVICGGETIKQOERAGVWVM--IQ 207

QY 317 ELTYLPERT-----ITVKNVETCAKAEIMKKIRE-----SYENDIASMNLQAHLP 366
Db 208 D-----GPQNTGADKPLRITGDPYKVOQAKEMVLELIRDQGGFREVNRVYSR-----IG 257

QY 367 GLNLNALGLFPPTSGMPPPTSGPPSANTPPYPQFEQSETEVHVQFIPALSVGAIIGKQG 426
Db 258 G-----NEGIDVP-----IPRFAVGIVIGRGE 280

QY 427 HIKOLSRFAGASIKIAPAEAPDAKVRMVIITGPPE-AQFKAQ----- 467
Db 281 MIKKIQNDAGVRIFQKPDGTTPE-RIAQITPPDRQCQHAAEIITDLRSVQAGNPGPGC 339

QY 468 --GRYVGK-----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNEL 513
Db 340 PGGRGGRGQGNMGMPPGGLOQEFNI-----VPTGKTGLIIGKGETIKSI 386

QY 514 QNLSSAEVVVPRDQTPDENDQV-VVKITGH---FYACOVAORKI 553
Db 387 SQSGGARIELQRPNNADPNMKLFTIRGTPQOIDYARQLIEKI 431

RESULT 14
US-09-261-855-23
Sequence 23, Application US/09261855A
Patent No. 6255055
GENERAL INFORMATION:
APPLICANT: Ross, Jeffrey
TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
FILE REFERENCE: 960296.95131
CURRENT APPLICATION NUMBER: US/09/261,855A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 47
TYPE: PRT
ORGANISM: Homo sapiens
US-09-261-855-23

Query Match 7.7%; Score 227; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 QFIPALSVGALIGKQGOHIKOLSRFAGASIKIAPAEAPDAKVRMVII 456
Db 1 QFIPALSVGALIGKQGOHIKOLSRFAGASIKIAPAEAPDAKVRMVII 47

RESULT 15
US-08-187-793-4
Sequence 4, Application US/08187793
Patent No. 5614371
GENERAL INFORMATION:
APPLICANT: POSNER, JEROME B.
APPLICANT: DARNELL, ROBERT B.
APPLICANT: FURNEAUX, HENRY M.
TITLE OF INVENTION: ANTIGEN RECOGNIZED BY ANTIBODIES ASSOCIATED
TITLE OF INVENTION: WITH PARANEOPLASTIC OPSOCLONUS AND METHODS OF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & DUNHAM
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/187,793
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 691,559
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39227
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: (212) 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-187-793-4

Query Match 7.7%; Score 227; DB 1; Length 530;
Best Local Similarity 22.4%; Pred. No. 5.5e-12;
Matches 119; Conservative 77; Mismatches 179; Indels 156; Gaps 20;

QY 155 PDMAAQNPLOQPRGRRGLGQRGSSRGSGVSKQKPCDPLRLVPTQFVGAIGKE 214
Db 43 PD---SRKRPLEAP-----PEAGSTKRTNTGEGGVF-----LKVLIPSYAAGSIIGK 88

QY 215 GATIRNITKQTSKIDVHRKENA-----GAAEKSTIILSTPEGTSACKSILIMHKEAQ 269
Db 89 GOTIVOLQKETGATIKLSKLSKSKDFYPGTTTTERVCLIQGTVEALNAVHGFIAEKIREMPQ 148

QY 270 DIKFTTEIPL-----KILAHNNFVGR 290
Db 149 NVAKTEPVSILOPOTTVNPDRIKQTLPLSPSTTTTKSSPDMTTSRANQVKIIVPNSTAGL 208

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 20:28:48 ; Search time 63.81 Seconds
(without alignments)

691.194 Million cell updates/sec

Title: US-09-685-696-176

Perfect score: 2956

Sequence: 1 MNKLIGNLSENAFSDLES.....VKHQOQKALQSGPPQSRRK 579

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: piri:*
2: piri2:*
3: piri3:*
4: piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 418.5 | 14.2 | 854 | 2 T23837 | hypothetical prote |
| 2 | 243.5 | 8.2 | 768 | 2 T27855 | hypothetical prote |
| 3 | 240 | 8.1 | 644 | 2 A53184 | myc far upstream e |
| 4 | 236.5 | 8.0 | 568 | 2 T49962 | hypothetical prote |
| 5 | 235.5 | 8.0 | 621 | 2 D96554 | hypothetical prote |
| 6 | 231.5 | 7.8 | 313 | 2 T48439 | probable RNA-bind |
| 7 | 229.5 | 7.8 | 398 | 2 T41600 | probable pre-mRNA |
| 8 | 227 | 7.7 | 510 | 2 T38489 | onconeural ventral |
| 9 | 225.5 | 7.6 | 589 | 2 T19216 | hypothetical prote |
| 10 | 225.5 | 7.6 | 611 | 2 T19217 | hypothetical prote |
| 11 | 222 | 7.5 | 479 | 2 C86275 | hypothetical prote |
| 12 | 219.5 | 7.4 | 396 | 2 A11224 | hnRNP protein - Af |
| 13 | 213.5 | 7.2 | 680 | 2 T25832 | hypothetical prote |
| 14 | 209.5 | 7.1 | 356 | 2 S58529 | alpha-complex prot |
| 15 | 207.5 | 7.0 | 1268 | 2 A44125 | high density lipop |
| 16 | 204.5 | 6.9 | 846 | 2 T04533 | hypothetical prote |
| 17 | 204 | 6.9 | 413 | 2 S46109 | hnRNP complex prot |
| 18 | 202.5 | 6.9 | 365 | 2 S42471 | hnRNP protein E2 - |
| 19 | 202 | 6.8 | 362 | 2 S78515 | single-stranded nu |
| 20 | 202 | 6.8 | 397 | 2 T30168 | hypothetical prote |
| 21 | 199.5 | 6.7 | 1279 | 2 T41389 | rna binding protei |
| 22 | 195 | 6.6 | 649 | 2 E84614 | probable RNA-bind |
| 23 | 194.5 | 6.6 | 1270 | 2 S23464 | vigilin - chichen |
| 24 | 190 | 6.4 | 463 | 2 S41495 | dC stretch-binding |
| 25 | 190 | 6.4 | 464 | 2 S43363 | transformation upr |
| 26 | 190 | 6.4 | 464 | 2 A54143 | kappa-B motif-bind |
| 27 | 189.5 | 6.4 | 632 | 2 T02627 | hypothetical prote |
| 28 | 184.5 | 6.2 | 1220 | 2 T19117 | hypothetical prote |
| 29 | 183 | 6.2 | 1222 | 2 S56030 | SCP160 protein - y |

30 174.5 5.9 806 2 T13690 hypothetical prote
31 172 5.8 495 2 T04255 hypothetical prote
32 170.5 5.8 629 2 T05425 polyadenylate-bind
33 166 5.6 633 1 DNHUPA polyadenylate-bind
34 166 5.6 1283 2 T49692 related to SCP160
35 164.5 5.6 473 2 A54691 octamer-binding pr
36 161 5.4 389 2 C86460 hypothetical prote
37 160.5 5.4 471 2 G01211 54 kDa protein - h
38 160.5 5.4 471 2 S41768 splicing factor ho
39 160.5 5.4 636 2 T48718 poly(A) binding pr
40 156 5.3 633 1 DNKLPA polyadenylate-bind
41 152.5 5.2 381 2 S45766 hypothetical prote
42 152.5 5.2 638 2 S37085 polyadenylate-bind
43 152 5.1 623 2 T07933 polyadenylate-bind
44 150 5.1 344 2 S59043 spilling factor SR
45 149 5.0 228 2 S42933 hypothetical prote

ALIGNMENTS

RESULT 1

T23837

hypothetical protein M88.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23837

R:Sulston, J.

submitted to the EMBL Data Library, June 1994

A:Reference number: Z19806

A:Accession: T23837

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-854 <WIL>

A:Cross-references: EMBL:Z34802; PIDN:CAA84338.1; GSPDB:GN00021; CESP:M88.5

A:Experimental source: clone M88

C:Genetics:

A:Gene: CESP:M88.5

A:Map position: 3

A:Introns: 38/3; 96/1; 133/3; 255/3; 307/3; 395/2; 436/2; 492/3; 657/3; 759/1; 820/1

Query Match 14.2%; Score 418.5; DB 2; Length 854;
Best Local Similarity 24.1%; Pred. No. 4.5e-19;
Matches 160; Conservative 103; Mismatches 215; Indels 187; Gaps 27;

Qy 78 QRIRKLIQIRNPPHLOWEVLDSLLVQYGVVSECEQVNTDSETAVN-----VTYSK 129

Db 167 QOMQOQOOSNQSAHFMHQQLQAVQQ-----QAQQMHRLQAGAPINPQQQFMVPPPTMMQP 221

Qy 130 DQARQALDKLNGFLENFTLVAYITPDEMA--AQO--NPLQQPRGRRLGQGRSSRQSGSP 185

Db 222 QOMQQA-QOQQAQQHMQHIIHQHPQMQQAQQGYHHPHQNQHQQA-GQHQSHSQ 279

Qy 186 G-----SVSKQKP-----C-----DLPLRLVPTQFVGALIGKEGATIRNITKQ 224

Db 280 NNQHRNNHNSQSGHPHIPQNLMPRCMLKDWPIRCVVEGKYHAVIIGPNSGTIKDIASS 339

Qy 225 TQSKID---VHRKNA--GAAEKSTILSTPGTGAACKSILEIMHKEA--QDIKETEI 277

Db 340 TRCRVDFVNLSSKERTVLGNDRILTVHGVAEQATKAVARILDVTQSEAKDDVNVGADT 399

Qy 278 PLKILAHNNFVGRLLGKGRNLKTEQDPTKTTTSPLOE-----LT-----LYNPRT 326

Db 400 VLRMAHQQLCGRLGKAGSSIKEIMQKTGNTITVTKYIEPPGGISGLTANELLGLMERT 459

Qy 327 ITVKG-NVETCAKAEIEIMKKIRESYENDIASMNLQAHILPGLNMLALFFPTSGMPP- 384

Db 460 IMVRGPISEAVVQAEALISAKLKKCYESD-SQLRAQSMQCP-----MPPMM-MPPI 508

Qy 385 -PTSGPPSAMTPPY-----PQFESETETVHQ----- 410

Db 509 LPPGASSAVSAPHFIPTPVGMVQVHFASSQHLVHQNNNSFLQPGVLQIQGTTLNRQ 568

```

Db 439 -----QFGGCGAQAQAQGEVI-----VPRLSAGMIIGKGEMIKRLAAETGTQIQ 485

Qy 441 IAPAEAPADKAVRMVLIITGPPEAQFKAQRIYGIKE--ENFVSPKEEVKLEA----- 490
      | : : : | : | : : | : | : : | : | : : | : | : : | : | : :
Db 486 FKPDTPNSEDRIAVINGTRDQIYRATERITEIVNRAIKNNGAPQDRGSAGTVLPQGSIF 545

Qy 491 HIRVPSFRANGRVIGGGKTVNELQNLSAEV-VPRDTPDTPDNDQVVVKITGH-----FYA 545
      : : | : | : | : | : : : : : : : : : | : | : | : | : | :
Db 546 YMHVPAGKCGLVIGKGENIKQIETRTGATCGLAPAAEQKNE-DEKVFEEKGSLQLIHHA 604

Qy 546 CQVAQRKIQEILTQVK-----QHQQOKALQSGPPQS 576
      : : | : | : | : | : : | : | : | : | : | : | : | : | :
Db 605 SHLVRIKVEISFNTVPVPPLOGAGGQIQQQQAMFSAGTQN 644

RESULT 3
A53184
myc far upstream element-binding protein - human
N:Alternate names: FUSE-binding protein
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C:Accession: A53184
R:Duncan, R.; Bazar, L.; Michelotti, G.; Tomonaga, T.; Kruttsch, H.; Avigan, M.; Leve
Genes Dev. 8, 465-480, 1994
A:Title: A sequence-specific, single-strand binding protein activates the far upstrea
A:Reference number: A53184; MUID:94170991
A:Accession: A53184
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <DUN>
A:Cross-references: GB:U05040; NID:g460151; PIDN:AAA17976.1; PID:g460152
C:Keywords: DNA binding

```

```

Query Match      8.1%; Score 240; DB 2; Length 644;
Best Local Similarity 22.3%; Pred. No. 9.1e-08;
Matches 100; Conservative 75; Mismatches 154; Indels 120; Gaps 17;

Qy 161 QONPLQ---QPRGRGLGSGSSRQSGSVSKOKPCDPLRLLLVPTQFVGAIIGREGAT 217
   | : || : || : || : | : | : | : | : | : | : | : | : | : | : |
Db 63 QKRPLEDGDPDAKKVAPQNDSEFGTQLPPMHQOQSRSSVMTEYKVPDGMWGFIIGRGEQ 122
   | : || : || : || : | : | : | : | : | : | : | : | : | : | : |
Qy 218 IRNITKTQSKIDVHRKENAGAAEKSTITLSTEGTSAACKSILEIMHKEAQDIIKTFE-- 275
   | : || : || : || : | : | : | : | : | : | : | : | : | : | : |
Db 123 ISRTQOESGCKIOI-APDSGGLPERSCMLTGTPEVSQSAKRLDDOIVEKGRPAGFHHGD 181
   | : || : || : || : | : | : | : | : | : | : | : | : | : | : |
Qy 276 ---BIPKILAHNNFVGRLLTGKGRNUKKITEQDQTDTKITISPLQELTYLNPERT---- 327
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 182 GPGNAVOEIMTPASKAGLVGCKGETIKQLQERAGVKVMV---IQD-----GPNTGADKPL 235
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 328 TVKGNVETCAKAEIIMKKIRE-----SYENDIASMNLQAHLLPGLNLNALGFLPPTSGM 382
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 236 RITGDPPYKVOQAKEMWLELIRDQGFREVEYEGSR-----IGG-----NEGI 278
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 383 PPPTSGPPSAMPPTYPQFEQSETETHQFIPALSVGAIIGKQGHQIKQLSRFAGASIKIA 442
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 279 DVP-----LPRFVGVITGRNEMIKKIONDAGVRIQFK 312
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 443 PAEAPDAKVRMVIITGPPE-AQFKAQ-----GRYK-----473
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 313 PDDGTTPE-RIAQITGPPDCQHAAELITDLLRSVQAGNPGPGPGRGRGRGQGNWNG 371
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 474 ----IKREENVSPKEEVKLEAHTRVPSFAAGRVIGGGKTVNQLNLSAEVVVPRDQTP 529
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 372 PPGGLQEFNFI-----VPTGTGLIIGRGGGETIKSISOOSGARIELQRNPPPP 418
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 530 DENDQV-VVKITGH----FYACOVAORKI 553
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 419 NADPNMKLFTIRTPQOIDYARQLIEEKI 447
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```


[illegible]

C:Genetics:
A:Gene: CESP:C12D8.1a
A:Map position: 5
A:Introns: 7/3; 25/1; 159/3; 318/3; 513/2

Query Match 7.6%; Score 225.5; DB 2; Length 589;

Best Local Similarity 23.4%; Pred. No. 6.8e-07;

Matches 86; Conservative 72; Mismatches 157; Indels 53; Gaps 11;

QY 202 VPTQFVGAIGEGATIRNITKOTQSDIVHRKENAGAAEKSIITILSTPEGTSACKSTL 261

Db 54 IPESAVGIVIGRGSEIOGIAKACRVQMSDPADPSSGVRVMTLEGSRNVETAKHLIN 113

QY 262 EIMKEAODIKTEEIP-----LKLAHNFVGLIGRGNLKKIEQDITDK-ITISPL 315

Db 114 EVVAR-SQNPRQYGFPAQTIDIAIPNRCGLIIGSGDTIRQLQKSGCKMILVDQN 172

QY 316 QELTYNPERTITVKGNETCAKAEIIMKKIRESYENDIASMNLOAHLIPGLNLNALGL 375

Db 173 QSVS--DQSKPLRITGDPQKIELA-KQLVAEILNSGGDGGGLQMHHAGG----- 221

QY 376 FPPTSGMPPTSGPPSAMTPPYQFEQSETVHQFIPALSVGAIGQGOHIKOLSRFA 435

Db 222 -----GGGASA-----RGEVV---VPRSSVGLIIGKOGDTIKRLAMET 256

QY 436 GASIKIAPAEAPDAKVRMVIITGPPEAQFKAQRIYGIKEENFVS-----PKEEVKLE 489

Db 257 GTKIQKPDDBSTPERCAVINGTRDQIYRATERITELVKKSTMQQGGGNVAGAWVNE 316

QY 490 A---HIRVPSFAAGRVIGKGTVNLQNLSSAEVVPDQTPDENDQVVKITGHFYAC 546

Db 317 ASTFTYMSVPAACKGLVIGKGETIKQINSEGAHCELSRDPGTNADEKVFV-IKGKRAI 375

QY 547 QVAQRKIQ 554

Db 376 EHAHLIR 383

RESULT 10

T19217

hypothetical protein C12D8.1b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19217

R:McMurray, A.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19092

A:Accession: T19217

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-611 <WIL>

A:Cross-references: EMBL:Z73969; PIDN:CAA98233.1; GSPDB:GNO0023; CESP:C12D8.1b

A:Experimental source: clone C12D8

C:Genetics:

A:Gene: CESP:C12D8.1b

A:Map position: 5

A:Introns: 181/3; 340/3; 535/2

Query Match 7.6%; Score 225.5; DB 2; Length 611;

Best Local Similarity 23.4%; Pred. No. 7.2e-07;

Matches 86; Conservative 72; Mismatches 157; Indels 53; Gaps 11;

QY 202 VPTQFVGAIGEGATIRNITKOTQSDIVHRKENAGAAEKSIITILSTPEGTSACKSTL 261

Db 76 IPESAVGIVIGRGSEIOGIAKACRVQMSDPADPSSGVRVMTLEGSRNVETAKHLIN 135

QY 262 EIMKEAODIKTEEIP-----LKLAHNFVGLIGRGNLKKIEQDITDK-ITISPL 315

Db 136 EVVAR-SQNPRQYGFPAQTIDIAIPNRCGLIIGSGDTIRQLQKSGCKMILVDQN 194

QY 316 QELTYNPERTITVKGNETCAKAEIIMKKIRESYENDIASMNLOAHLIPGLNLNALGL 375

Db 195 QSVS--DQSKPLRITGDPQKIELA-KQLVAEILNSGGDGGGLQMHHAGG----- 243

QY 376 FPPTSGMPPTSGPPSAMTPPYQFEQSETVHQFIPALSVGAIGQGOHIKOLSRFA 435

Db 244 -----GGGASA-----RGEVV---VPRSSVGLIIGKOGDTIKRLAMET 278

QY 436 GASIKIAPAEAPDAKVRMVIITGPPEAQFKAQRIYGIKEENFVS-----PKEEVKLE 489

Db 279 GTKIQKPDDBSTPERCAVINGTRDQIYRATERITELVKKSTMQQGGGNVAGAWVNE 338

QY 490 A---HIRVPSFAAGRVIGKGTVNLQNLSSAEVVPDQTPDENDQVVKITGHFYAC 546

Db 339 ASTFTYMSVPAACKGLVIGKGETIKQINSEGAHCELSRDPGTNADEKVFV-IKGKRAI 397

QY 547 QVAQRKIQ 554

Db 398 EHAHLIR 405

RESULT 11

C86275

hypothetical protein AAD39302.1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86275

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86275

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <STO>

A:Cross-references: GB:AE005172; NID:g5080792; PIDN:AAD39302.1; GSPDB:GNO0141

C:Genetics:

A:Map position: 1

Query Match 7.5%; Score 222; DB 2; Length 479;

Best Local Similarity 21.6%; Pred. No. 8.4e-07;

Matches 98; Conservative 93; Mismatches 169; Indels 94; Gaps 19;

QY 165 LQOPRGRRG-----LGQGRSSRQSGSPGVSQKQPC--DLPLRLLVPTQFVGAIIKEGA 216

Db 5 LRNIHGKRNLOSEFTGNGSGSKRRNLHDETDQNVIASEDVTVRYLCPVKTKGIIGKGE 64

QY 217 TIRNITKOTQSDIVHRKENAGAAEKSIITILSTPEGTS-----AACKSILEIMHK--- 266

Db 65 IAKQIRSETKSNMRIN-EALPGCEERVVMTYSTNEELNHFGDDGELVCPA-LDALFKVHD 122

QY 267 -----EAQDIKFTTEIPLKILAHNFVGRILGKGRNLKKIEQDITDKITI- 312

Db 123 MVVADADQDDGTDNDNDLGEKQTVTVRMLVPSDQICGVIGKGGQVTONLRNDTNAQIRVI 182

QY 313 ---SPLQELTYNPERTITVKGNETCAKAEIIMKKIRES---YENDIASMNLOAHLIP 366

Db 183 KHLFPACALTLSHDELLII-GEPLVVRALYQVVALSHLDNPSRFQHLILLSSSSSSHHQP 241

QY 367 GLNLNALGLFPPTSGMPTSGPPSAMTPPYQFE-QSETETVHQF-----IPALSVGAI 420

Db 242 GAMLMS-----AALTSSHRNYAVRDIADAREFCVCFICPAENVGGV 283

QY 421 IKGQGOHIKOLSRFAGASKIAPAEAPDAKVRMVIITGPPE-----AQFKAGRI 470

Db 284 IKGGGGFNIQRQETGATIRVNTSETDDDD--CIIFISSKEFYEDQSPAVNAIRLQORC 341

QY 471 YGKI-KEENFYPKKEVKLEAHIRVPSFAAGRVICKGKTKYNELONLSSAEV-VVPRDQT 528
Db 342 SEKVGDKN-----DLAISTRLLVSSQIGCLIGKGGAVISEMRSVTRNIRILOKEDV 395
QY 529 PD--ENDQVVVKITGHFVACQAKKIQEILTQV 560
Db 396 PKIAREDEEMVQITGSPDAAMKA-----LTQV 422

RESULT 12
S41224
hnRNP protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S41224
R:Strom, H.; Matunis, M.J.; Michael, W.M.; Dreyfuss, G.
Nucleic Acids Res. 21, 1193-1198, 1993
A:Title: The pre-mRNA binding K protein contains a novel evolutionarily conserved motif.
A:Reference number: S41224; MUID:93219080
A:Accession: S41224
A:Molecule type: mRNA
A:Residues: 1-396 <STO>
C:Superfamily: transformation upregulated nuclear protein HNRPK

Query Match 7.4%; Score 219.5; DB 2; Length 396;
Best Local Similarity 21.3%; Pred. No. 9.2e-07;
Matches 93; Conservative 76; Mismatches 166; Indels 101; Gaps 18;

QY 175 GORSSRQSGSPGVSQKPCDL--PLRLVPTQFVGAIKKEGATINIKQTSKIDVH 232
Db 17 GKRAEDMEDEQAFKRNRTDMVLEKILLOSKNAGAVIGKGNKIKALRTDYNASVSV- 75
QY 233 KENAGAAREKSITILSTPETSAAKSSILEIMHKEAQDIKFTTEIPKILAHNNFVGRLI 292
Db 76 --PDSSGPERILSADIEITGELKKIPTLEHFKGNDF--DCELRLLIHSLAGGII 131
QY 293 GKEGRNLKKEQDQTDKITISPLQELTYNPRTITVKNVETCAKAEIEIMKKIRES-- 350
Db 132 GVRGAKIKELREKTQT--TIKFECCPHSTDRVLIGRPRVVECIKVILDLISESPV 189
QY 351 -----YEN-DIASNNL-----QAHLPGLNALGLF-----PPTSGMPPTTS 387
Db 190 KGRSQYDPNFYETDYGGMFDDRRGRPH---GFSMHARGFDRMPGCGGRMPQS 246
QY 388 -----GPPSAMTPPYPOFEQSETETVHQFIPALSVGAIIGKQGHKQLSRFAGAS-- 438
Db 247 RRDYDDMSPRRGLPLPPP-----GRGR-----GGSRA 274
QY 439 --TKIAPAEAPDAKVRMVIITGPPEAQFKAQRIYKIKKEENFVSPKEVKLEAHIRVPS 496
Db 275 RNLPLPPPPPPRGDGR-----RGRPDHYDGMGRGGRGSGFDIGGP-----VITQTVPK 326
QY 497 FAAGRVIGKGVKNELONLSSAEVVPDQTPDENDOVVVKITGHFYACQVAKKIQE- 555
Db 327 DLAGSIIGKGGKQIKQIRHESGASIKI--DEPLEGSDRIITITG-----TQDQIQNA 377
QY 556 ---ILTVQKHQHOQKA 568
Db 378 RFLQNSVVKQFSEDYA 393

RESULT 13
T25832
hypothetical protein M01A10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25832
R:Scheet, P.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid M01A10.
A:Reference number: 220094

A:Accession: T25832
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-680 <SCH>
A:Cross-references: EMBL:U88174; PIDN:AA842272.1; GSPDB:GN00019; CESP:M01A10.1
A:Experimental source: strain Bristol N2; clone M01A10
C:Genetics:
A:Gene: CESP:M01A10.1
A:Map position: 1
A:Introns: 63/2; 97/3; 121/1; 160/1; 269/3; 411/3; 512/1; 649/2

Query Match 7.2%; Score 213.5; DB 2; Length 680;
Best Local Similarity 21.3%; Pred. No. 5e-06;
Matches 73; Conservative 69; Mismatches 139; Indels 61; Gaps 9;

QY 202 VPTQFVGAIKKEGATINIKQTSKIDVHRKENAGAAEKSIITILSTPETSACKSIL 261
Db 152 IPESVGLVIGRNGVEIQAISQKSCRVQI-VAEPSTGYRSVDIYGISENIEVAKKLIN 210
QY 262 EIM---HKEAQD-----IKFTEIPLKILAHNNFVGRILGKGRNLKKEQ 304
Db 211 EVVARGKLSQELPCSVPOFQPIPAVSNSSKVTIIIPIPANKCGAIGKKEQMRKLS 270
QY 305 DTDKTIISPLQELTYNPRTITVKNVETCAKAEIEIMKKIRESYENDIASMNLQHL 364
Db 271 WTNCDFIL--IQENNIADSVKPLQITGPKVEHAHA-----LVADI 310
QY 365 IGPLNALGLFPPTSGMPPTSGPPSAMTPPYPOFEQSETETVHQFIPALSVCAIGKQ 424
Db 311 LDGFD-----ECPPAGMAGNSPVAAMS-----LQVKVPRSTVGAINGLQ 349
QY 425 GQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQRIYKIKKEENFVSPKE 484
Db 350 GSNIKKTSNETKIQMPDDPKLMERTLVVIGNKKNVYVC-ARLQKIVEANSENANT 408
QY 485 EVKLEAHIRVPSFAAGRVIGKGVKNELONLSSAEVVPD 526
Db 409 PISL-FYMLIPASKGLVIGRGTETIRQINKSGAYCEMSRD 449

RESULT 14
S58529
alpha-complex protein 1 - human
N:Alternate names: nucleic acid-binding protein; protein PCBP-1
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C:Accession: S58529; S58523; S65678; S43489; S41378; S42472
R:Kiledjian, M.; Wang, X.; Liehaber, S.A.
EMBO J. 14, 4357-4364, 1995
A:Title: Identification of two KH domain proteins in the alpha-globin mRNA stability
A:Reference number: S58523; MUID:96016208
A:Accession: S58529
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-356 <KIL>
A:Accession: S58523
A>Status: preliminary
A:Molecule type: protein
A:Residues: 125-139;251-265;315-323 <K12>
R:Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A:Title: Characterisation of two major cellular poly(rC)-binding human proteins, each
A:Reference number: S65678; MUID:95331278
A:Accession: S65678
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-204,'v',206-356 <LEF>
A:Cross-references: EMBL:X78137; NID:g460770; PIDN:CAA55016.1; PID:g460771
A:Experimental source: AMA cells (transformed human amnion cells)
A:Note: submitted to the EMBL Data Library, March 1994
R:Aasheim, H.C.; Loukianova, T.; Deggerdal, A.; Smeland, E.B.
Nucleic Acids Res. 22, 959-964, 1994

A:Title: Tissue specific expression and cDNA structure of a human transcript encoding a
A:Reference number: S43489; MUID:94203810
A:Accession: S43489
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-298, 'H' <AAS>
A:Cross-references: EMBL:D29505; NID:9444020; PIDN:CAA82631.1; PID:9444021
C:Genetics:
A:Gene: GDB:HNRNPX
A:Cross-references: GDB:344947
C:Keywords: RNA binding

Query Match 7.1%; Score 209.5; DB 2; Length 356;
Best Local Similarity 23.4%; Pred. No. 3.5e-06;
Matches 83; Conservative 63; Mismatches 118; Indels 91; Gaps 12;

Qy 196 LPLRLVPTQVGAIGKEGATIRNITKOTSKIDVHRKENAGAEEKSITILSTPEGTSA 255
Db 14 LTIRLLMHGKEVGIIGKKGESVKRIEESGARINI-----SEGNCPERIITLTGP--TNA 67
Qy 256 ACKSILEIMHKEAQDIKFT-----EEIPLKILAHNFEVRLIGKEGRNLKKIEODT 306
Db 68 IFKAFAMIIDLEEDINSMTNSTAASPPVTLRLVVPATQCGSLIGKGGCKIKIREST 127
Qy 307 DTKITISPLQELTYNPERTITVKGNETCAKAEIEIMKKIRES-----YE- 352
Db 128 GAQVQVA--GDMLPNSTERAITAGVPQSVTECVKQICLVMLETLSQSPQGRVMTIPYQP 185
Qy 353 -----NDIASMNLQAHILPG-----LNLNALGL----- 375
Db 186 MPASSPVICAGGQDCRDAAGYPHATDLEGPLDAYSIQGHITISPLDLAKLNOVARQQ 245
Qy 376 -----FPPTSGMPPTSGPPSAMPYPQFEQSEETVTHQF-IPALSVGAIIGK 423
Db 246 SHFAMHGGTGFAGIDSSPEVKG-----YNASLDASTQTHLTPNNLIGCIIGR 297
Qy 424 QGQHIKQLSRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQRIYGRKKEE 477
Db 298 QGANINEIRQWGAQIKIANPVEGSSG--RQVITGSAASISLAQYLINARLSSE 350

RESULT 15

A44125
high density lipoprotein-binding protein, 110K - human
C:Species: Homo sapiens (man)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C:Accession: A44125
R:McKnight, G.L.; Reasoner, J.; Gilbert, T.; Sundquist, K.O.; Hokland, B.; McKernan, P.A.
J. Biol. Chem. 267, 12131-12141, 1992
A:Title: Cloning and expression of a cellular high density lipoprotein-binding protein
A:Reference number: A44125; MUID:92291094
A:Accession: A44125
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <MCK>
A:Cross-references: GB:M64098; GB:M83789; NID:g183891; PIDN:AAA35962.1; PID:g183892
A:Note: sequence extracted from NCBI backbone (NCBIN:106862, NCBIIP:106863)
C:Superfamily: vigillin

Query Match 7.0%; Score 207.5; DB 2; Length 1268;
Best Local Similarity 23.9%; Pred. No. 3e-05;
Matches 95; Conservative 69; Mismatches 153; Indels 81; Gaps 18;

Qy 202 VPTQFVGAIGKEGATIRNITKOTSKIDVHRKENAGAEEKSITILSTPEGTSAACKSIL 261
Db 157 IPKEHHRVICKNGEKLDLEKLTATKIPIRPDD---PSNQIKITGPKGIEKARHEVL 213
Qy 262 EIMHKEAQDIKFTTEIPLKILAHNFVG---RLIGKEGRNLKKIEQDTHDKITISPLQEL 318
Db 214 LI--SAEQDKRAVERLEVEKAFHFPIAGPNRLVGV-----EIMQETGRINIPP---- 260

Qy 319 TLYNPERT-ITVKGNETCAKAEIEIMKKIRESYENDIASMNL-----QAHILPGLNLNA 372
Db 261 --PSVNRTEIVFTGEKEQLAQAVARI-KKIYBEKKKKTITIAVEVKKSQHKVYIGPKGNS 317
Qy 373 LGLFPPTSG-----MPPPTS-----GPP-----SAMTPPYQFEQSEETV-----HQ 410
Db 318 LOEILERTGVSVSEIIPPSDSISSETVILRGPEKLQALTEVYAKANSFTVSSVAAPSWLHR 377
Qy 411 FIPALSVGAIIIGKQGOHIKQLSRFAGASIKIAPAPADAKVRMVIITGPPEAQFKAQGR I 470
Db 378 F-----IIGKQGNLAKITQ-QMPKVHIEFTEGED-----KITLEGPTEDVNVAAEQEI 424
Qy 471 YGKIKE-----ENFVSPKBEVKLEAHIRVPSFAAGRVIGKGTVNELQNLSSAEVVVPRD 526
Db 425 EGMVADLINRMDYVEINIDHKFHRHL-----IGKSGANINRIKQYKVSVRIP-- 472
Qy 527 QTPDENQVVKITGHFYACQVAQKIQEILLTQVKOHQ 564
Db 473 --PDSEKSNLIRIEGDPQGVQQAQKRELLELASRMENER 508

Search completed: January 25, 2002, 22:17:32
Job time: 6524 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 22:12:48 ; Search time 52.07 Seconds
(without alignments)
407.700 Million cell updates/sec

Title: US-09-685-696-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAPSLES.....VKHQHQKALQSGPPQSRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|------------|---------------------|
| 1 | 227 | 7.7 | 510 | 1 | NOA1_HUMAN | P51513 homo sapien |
| 2 | 224.5 | 7.6 | 339 | 1 | PCB3_HUMAN | P57721 homo sapien |
| 3 | 219.5 | 7.4 | 339 | 1 | PCB3_MOUSE | P57722 mus musculus |
| 4 | 213.5 | 7.2 | 403 | 1 | PCB4_HUMAN | P57723 homo sapien |
| 5 | 212.5 | 7.2 | 403 | 1 | PCB4_MOUSE | P57724 mus musculus |
| 6 | 210.5 | 7.1 | 606 | 1 | Y475_ARATH | P58223 arabidopsis |
| 7 | 209.5 | 7.1 | 356 | 1 | PCB1_HUMAN | Q15365 homo sapien |
| 8 | 209.5 | 7.1 | 356 | 1 | PCB1_RABIT | Q19048 oryctolagus |
| 9 | 207.5 | 7.0 | 1268 | 1 | VGLN_HUMAN | Q00341 homo sapien |
| 10 | 204 | 6.9 | 413 | 1 | PBP2_YEAST | P38151 saccharomyc |
| 11 | 202.5 | 6.9 | 365 | 1 | PCB2_HUMAN | Q15366 homo sapien |
| 12 | 202 | 6.8 | 362 | 1 | PCB2_MOUSE | Q61990 mus musculus |
| 13 | 194.5 | 6.6 | 1270 | 1 | VGLN_CHICK | P81021 gallus gall |
| 14 | 190 | 6.4 | 463 | 1 | ROK_HUMAN | Q07244 homo sapien |
| 15 | 190 | 6.4 | 463 | 1 | ROK_RABIT | Q19049 oryctolagus |
| 16 | 190 | 6.4 | 464 | 1 | ROK_MOUSE | Q60577 mus musculus |
| 17 | 183 | 6.2 | 1222 | 1 | S160_YEAST | P06105 saccharomyc |
| 18 | 170.5 | 5.8 | 629 | 1 | PAB2_ARATH | P42731 arabidopsis |
| 19 | 164.5 | 5.6 | 644 | 1 | PAB4_HUMAN | Q13310 homo sapien |
| 20 | 160.5 | 5.4 | 636 | 1 | PAB1_MOUSE | P29341 mus musculus |
| 21 | 159.5 | 5.4 | 470 | 1 | NR54_HUMAN | Q15233 homo sapien |
| 22 | 159.5 | 5.4 | 636 | 1 | PAB1_HUMAN | P11940 homo sapien |
| 23 | 156 | 5.3 | 633 | 1 | PABP_XENLA | P20965 xenopus lae |
| 24 | 152.5 | 5.2 | 381 | 1 | YBP2_YEAST | P38199 saccharomyc |
| 25 | 150 | 5.1 | 344 | 1 | SFR6_HUMAN | Q13247 homo sapien |
| 26 | 147.5 | 5.0 | 747 | 1 | FCA_ARATH | Q04425 arabidopsis |
| 27 | 145.5 | 4.9 | 414 | 1 | NSR1_YEAST | P27476 saccharomyc |
| 28 | 145 | 4.9 | 414 | 1 | NOP3_YEAST | Q01560 saccharomyc |
| 29 | 139.5 | 4.7 | 522 | 1 | PAB2_HUMAN | Q15097 homo sapien |
| 30 | 138.5 | 4.7 | 424 | 1 | S3B4_HUMAN | Q15427 homo sapien |
| 31 | 135.5 | 4.6 | 359 | 1 | ELV2_HUMAN | Q12926 homo sapien |
| 32 | 135.5 | 4.6 | 360 | 1 | ELV2_MOUSE | Q60899 mus musculus |
| 33 | 135 | 4.6 | 359 | 1 | ELV3_HUMAN | Q14576 homo sapien |

| | | | | | | |
|----|-------|-----|------|---|------------|---------------------|
| 34 | 135 | 4.6 | 367 | 1 | ELV3_MOUSE | Q60900 mus musculus |
| 35 | 135 | 4.6 | 494 | 1 | SFR4_HUMAN | Q08170 homo sapien |
| 36 | 134 | 4.5 | 285 | 1 | CABA_MOUSE | Q99020 mus musculus |
| 37 | 134 | 4.5 | 307 | 1 | SXL_MEGSC | O01671 megaselia s |
| 38 | 133 | 4.5 | 373 | 1 | ELV4_RAT | O09032 rattus norv |
| 39 | 133 | 4.5 | 380 | 1 | ELV4_HUMAN | P26378 homo sapien |
| 40 | 133 | 4.5 | 385 | 1 | ELV4_MOUSE | Q61701 mus musculus |
| 41 | 130 | 4.4 | 633 | 1 | ROR_HUMAN | Q43390 homo sapien |
| 42 | 129.5 | 4.4 | 6359 | 1 | BACC_BACLI | O68008 b bacitraci |
| 43 | 129 | 4.4 | 653 | 1 | PABP_SCHPO | P31209 schizosacch |
| 44 | 128.5 | 4.3 | 660 | 1 | PAB3_ARATH | O64380 arabidopsis |
| 45 | 128 | 4.3 | 576 | 1 | PABP_YEAST | P04147 saccharomyc |

ALIGNMENTS

| | | | | |
|----------|--|-----------|------|---------|
| RESULT 1 | NOA1_HUMAN | STANDARD; | PRT; | 510 AA. |
| ID | NOA1_HUMAN | | | |
| AC | P51513; | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | |
| DE | ONCONEURAL VENTRAL ANTIGEN-1 (NOVA-1) (PARANEOPLASTIC RI ANTIGEN) | | | |
| DE | (VENTRAL NEURON-SPECIFIC PROTEIN 1). | | | |
| GN | NOVA1. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=Cerebellum, and Hippocampus; | | | |
| RX | MEDLINE=94000830; PubMed=8398153; | | | |
| RA | Buckanovich R.J., Posner J.B., Darnell R.B.; | | | |
| RT | "Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding | | | |
| RT | protein and is specifically expressed in the developing motor | | | |
| RT | system."; | | | |
| RL | Neuron 11:657-672(1993). | | | |
| RP | [2] | | | |
| RP | SEQUENCE OF 1-34 FROM N.A. | | | |
| RC | TISSUE=Fetal brain; | | | |
| RA | Dmitrenko V.V., Garifulin O.M., Shostak K.A., Smikodub A.I., | | | |
| RA | Kavsan V.M.; | | | |
| RL | Submitted (APR-1996) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -!- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC | | | |
| CC | SUBSET OF DEVELOPING NEURONS. | | | |
| CC | -!- SUBCELLULAR LOCATION: NUCLEAR. | | | |
| CC | -!- TISSUE SPECIFICITY: BRAIN. | | | |
| CC | -!- DISEASE: ANTIBODIES ARE SEEN AGAINST THIS ANTIGEN IN THE PATIENTS | | | |
| CC | SUFFERING WITH PARANEOPLASTIC OPSOCLONUS-ATAXIA (POA). POA IS A | | | |
| CC | DISORDER IN WHICH ABNORMAL MOTOR CONTROL OF THE EYES, TRUNK AND | | | |
| CC | LIMBS DEVELOP IN WOMEN WITH BREAST OR SMALL LUNG CANCER. POA | | | |
| CC | PATIENTS SUFFER FROM OPSOCLONUS, A CHAOTIC EYE MOVEMENT DISORDER | | | |
| CC | ATTRIBUTABLE TO A LACK OF INHIBITION OF BURST NEURONS IN THE BRAIN | | | |
| CC | STEM: MYOCLONUS, A DISORDER PERHAPS ATTRIBUTING TO A FAILURE OF | | | |
| CC | INHIBITORY CONTROL OVER SPINAL MOTOR NEURONS; AND TRUNCAL ATAXIA, | | | |
| CC | REFERABLE TO CEREBELLAR DYSFUNCTION; THERE ARE NO CORTICAL OR | | | |
| CC | SENSORY DEFICITS. | | | |
| CC | -!- SIMILARITY: CONTAINS 3 KH DOMAINS. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; U04840; AAA16022.1; | | | |
| DR | EMBL; Z70771; CAA94810.1; | | | |
| DR | MIM; 602157; -. | | | |


```

RX MEDLINE=20396135; PubMed=10936052;
RA Makeyev A.V., Liebhauer S.A.;
RT "Identification of two novel mammalian genes establishes a subfamily
RL of KH-domain RNA-binding proteins.";
RL Genomics 67:301-316(2000)
CC -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF176327; AAC09238.1; -
DR MGD: MGI:1890470; Pcbp3.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PS00084; KH_TYPE_1; 3.
DR Nucleic acid protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW Repeat.
FT DOMAIN 13 63 KH 1.
FT DOMAIN 97 150 KH 2.
FT DOMAIN 261 313 KH 3.
FT DOMAIN 339 399 AA; A6F1C7C176A64F9C CRC64;
SQ SEQUENCE 339 AA; 35958 MW; A6F1C7C176A64F9C CRC64;

Query Match 7.4%; Score 219.5; DB 1; Length 339;
Best Local Similarity 20.6%; Pred. No. 1e-06; Indels 129; Gaps 7;
Matches 75; Conservative 52; Mismatches 108;

QY 277 IPLKILAHNNFVGRIGRNLKKEIDTDTKITISPLQELTYNPTITVKGNVETC 336
D 14 LTIRLLMHGKEVSGIIGKKEVTKMKRESGARINISEGN-----CPEIVITITGPTDAI 68
QY 337 AKAEIEIMKKIKRESYENDIASNNLQAHLPGLNGLNGLFPPTSGMPPPTSGPPSAMTPP 396
D 69 KFAFIAMAYKFEEDILNSMN-----SPATSKPP----- 97
QY 397 YPQFEQSETEVHQFIPALSVGAIIGKOGHITKQLSRFAGASIKIAPAEAPDAKVMVII 456
D 98 -----VTLRLVVPASQCSGLIGKGGSKIKREISTGAQVQVAGOMLPNSTERAVTI 148
QY 457 TGPPEAQFK-----AQGRIV----- 471
D 149 SCTPDIAFCQVKQICVVMLESPPKGTATPYRKPACTPVIFAGGQAYTIQOYAIHPDQ 208
QY 472 -----GKIEENFVSPKEVKLEA-----HIRV 494
D 209 LTKHLQAMQOTPPFPGLQTNPAFPCKELPLHSSEEAQNLMSQSGLDASPPASTHETLI 268
QY 495 PSFAAGRVIGKGGKVNELQNLSSREVVPVPRDTPDENDQVVKVITGHEYACQVQRKIQ 554
D 269 PNDLIGCLIGROGTKINEIROMSGAQIKIANTEGSSERQIT--ITGTPANISLAQYLIN 326
QY 555 EILT 558
D 327 ARLT 330

RESULT 4
PCB4_HUMAN
ID PCB4_HUMAN STANDARD; PRT; 403 AA.
AC P57723.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Created)

```

```

DE POLY(RC)-BINDING PROTEIN 4 (ALPHA-CP4).
GN PCBp4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20396135; PubMed=10936052;
RA Makeyev A.V., Liebhauer S.A.;
RT "Identification of two novel mammalian genes establishes a subfamily
RL of KH-domain RNA-binding proteins.";
RL Genomics 67:301-316(2000)
CC -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF176330; AAG09241.1; -
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PS00084; KH_TYPE_1; 3.
DR Nucleic acid protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW Repeat.
FT DOMAIN 17 67 KH 1.
FT DOMAIN 101 154 KH 2.
FT DOMAIN 241 293 KH 3.
FT SEQUENCE 403 AA; 41481 MW; 3D99F762A9471265 CRC64;
SQ SEQUENCE 403 AA; 41481 MW; 3D99F762A9471265 CRC64;

Query Match 7.2%; Score 213.5; DB 1; Length 403;
Best Local Similarity 25.6%; Pred. No. 3.1e-06; Indels 49; Gaps 13;
Matches 81; Conservative 61; Mismatches 126;

QY 183 GSPGSVSKQKPCD--LPLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKENAGAA 240
D 3 GSDGGLLEEPESLITLRLMLHGKEVSGIIGKKEVTKMKRESGARITI-----SEGSC 58
QY 241 EKSITILSTPETSACKSILEIMHKEAQDI-----KFTTEIPLKILAHNNFVGRIL 291
D 59 PERITIT--GSTAAVFHAVSMIAFKLDEDLCAAPANGNVSRPPVTLRLVIPASQCSGL 116
QY 292 IGKEGRNLKKIEQDQDTTKITISPLQELTYNPTITVKGNVETCAKAEIEIMKKIKRESY 351
D 117 IGKAGTKIKIEIRETTGAQVQA--GDLPLNSTERAVTISGVDPDAIILCVRQICAVILESP 174
QY 352 ENDIASNMQLAHLPGLNGLNGLFPPTSGMPPPTSGPPSAMTP-----PY 397
D 175 PK---GATIPYH--PSLSLGTV--LLSANQGF--SVQGYGAVTPAEVTKLQQLSSHAVPF 226
QY 398 -----PQFEQ--SETETVHQFIPALSVGAIIGKOGHITKQLSRFAGASIKIAPAEAPDAK 450
D 227 ATPSVVPGLDPGTQTSSQEFVLPNDLIGCVIGRQSGKISIRQMSGAHIKIK--NQAEAG 285
QY 451 VRMVIITGPPPEAQFKAQ 467
D 286 ERHVTITGSPVSIALAQ 302

RESULT 5
PCB4_MOUSE
ID PCB4_MOUSE STANDARD; PRT; 403 AA.
AC P57724;
DT 20-AUG-2001 (Rel. 40, Created)

```


Db 68 IFKAFAMIDKLEEDINSSMTNSTAASRPVTLRLVVDPATQCGSLGKGCKIKEIREST 127
QY 307 DTKITISPLQELTYLNPRTITVKGNETCAKAEETMKKIKRES-----YE- 352
Db 128 GAQVOVA--GDMLPNSTERAITAGVPSQVTECVKQICLVNLETLSQSGRVMVTIPYQP 185
QY 353 -----NDIASMNIQAHLIPG-----LNLNALGL----- 375
Db 186 MPASSPVICAGGQDRCSDAVGYPHATHDLEGPPLDAYSIQOHTISPDLAKLNOVARQ 245
QY 376 -----FPPTSCMPPTSGPPSAMPYPQFQSETEVHQF-IPALSVAIGK 423
Db 246 SHFAMHGGTGAGIDSSSEVKG-----YWASLDASTQTTHETIPNNLIGCIIGR 297
QY 424 QGOHIKOLSRFAGASIKIA-PAPAPDAKVRMVIITGPPEAQFKAQRIYKIKEE 477
Db 298 QGANINEIROMSQAQIKIANPVEGSG--RQVITGSAASISLAQYLINARLSSE 350

RESULT 8

PCB1_RABIT
ID PCB1_RABIT STANDARD; PRT; 356 AA.
AC O19048;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POLY(RC)-BINDING PROTEIN 1 (ALPHA-CP1) (HNRNP-E1).
GN PCB1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99207122; PubMed=10101190;
RA Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT "Tissue-specific translational regulation of alternative rabbit
RT 15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL Nucleic Acids Res. 27:1828-1836(1999).
CC -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT BINDS
CC PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ003023; CA005814.1; .
CC InterPro: IPR000958; KH.
CC Pfam; PF00013; KH-domain; 3.
CC SMART; SM00322; KH; 3.
CC PROSITE; PS50084; KH_TYPE_1; 3.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding; Repeat.
FT DOMAIN 13 75 KH 1.
FT DOMAIN 97 162 KH 2.
FT DOMAIN 279 343 KH 3.
SQ SEQUENCE 356 AA; 37497 MW; 6D1A261276CA206D CRC64;

Query Match 7.1%; Score 209.5; DB 1; Length 356;
Best Local Similarity 23.4%; Pred. No. 4.5e-06;
Matches 83; Conservative 63; Mismatches 118; Indels 91; Gaps 12;
QY 196 LPLRLVPTQFVGAIKKEGATINITKQTSKIDVHRKENAGAEKSIITLSTPGETSA 255
Db 14 LTIRLLMHGKGVGSIIGKSGVKRIRESGARINI-----SEGCNCPRIITLGP--TNA 67
QY 256 AKSILEIMHKEAQDIKFT-----EIPKLIAHNNFVRLICGKGRNLKKIEQDT 306

Db 68 IFKAFAMIDKLEEDINSSMTNSTAASRPVTLRLVVDPATQCGSLGKGCKIKEIREST 127
QY 307 DTKITISPLQELTYLNPRTITVKGNETCAKAEETMKKIKRES-----YE- 352
Db 128 GAQVOVA--GDMLPNSTERAITAGVPSQVTECVKQICLVNLETLSQSGRVMVTIPYQP 185
QY 353 -----NDIASMNIQAHLIPG-----LNLNALGL----- 375
Db 186 MPASSPVICAGGQDRCSDAVGYPHATHDLEGPPLDAYSIQOHTISPDLAKLNOVARQ 245
QY 376 -----FPPTSCMPPTSGPPSAMPYPQFQSETEVHQF-IPALSVAIGK 423
Db 246 SHFAMHGGTGAGIDSSSEVKG-----YWASLDASTQTTHETIPNNLIGCIIGR 297
QY 424 QGOHIKOLSRFAGASIKIA-PAPAPDAKVRMVIITGPPEAQFKAQRIYKIKEE 477
Db 298 QGANINEIROMSQAQIKIANPVEGSG--RQVITGSAASISLAQYLINARLSSE 350

RESULT 9
VGLN_HUMAN
ID VGLN_HUMAN STANDARD; PRT; 1268 AA.
AC Q00341; Q9UCY3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VIGILIN (HIGH DENSITY LIPOPROTEIN-BINDING PROTEIN) (HDL-BINDING
DE PROTEIN).
GN HDLBP OR HBP OR VGL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92291094; PubMed=1318310;
RA McKnight G.L., Reasoner J., Gilbert T., Sundquist K.O., Hokland B.,
RA McKernan P.A., Champagne J., Johnson C.J., Bailey M.C., Holly R.,
RA O'Hara P.J., Oram J.F.;
RT "Cloning and expression of a cellular high density lipoprotein-binding
RT protein that is up-regulated by cholesterol loading of cells.";
RL J. Biol. Chem. 267:12131-12141(1992).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=96184515; PubMed=8605996;
RA Kugler S., Grunweller A., Probst C., Klinger M., Muller P.K.,
RA Kruse C.;
RT "Vigilin contains a functional nuclear localisation sequence and is
RT present in both the cytoplasm and the nucleus.";
RL FEBS Lett. 382:330-334(1996).
CC -!- FUNCTION: APPEARS TO PLAY A ROLE IN CELL STEROL METABOLISM.
CC IT MAY FUNCTION TO PROTECT CELLS FROM OVER-ACCUMULATION OF
CC CHOLESTEROL.
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 14 KH DOMAINS.
CC -!- SIMILARITY: STRONG, TO C.ELEGANS C08H9.2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M64098; AAA35962.1; .
CC MIM; 142695; .
CC InterPro: IPR000958; KH.
CC Pfam; PF00013; KH-domain; 14.
CC SMART; SM00322; KH; 14.
CC PROSITE; PS50084; KH_TYPE_1; 14.
KW HDL; Lipid transport; Cholesterol metabolism; RNA-binding; Repeat;

```
KW Phosphorylation; Nuclear protein.
FT DOMAIN 158 229 KH 1.
FT DOMAIN 230 230 KH 2.
FT DOMAIN 303 371 KH 3.
FT DOMAIN 372 442 KH 4.
FT DOMAIN 443 514 KH 5.
FT DOMAIN 515 588 KH 6.
FT DOMAIN 589 660 KH 7.
FT DOMAIN 661 734 KH 8.
FT DOMAIN 735 807 KH 9.
FT DOMAIN 808 880 KH 10.
FT DOMAIN 881 979 KH 11.
FT DOMAIN 980 1059 KH 12.
FT DOMAIN 1060 1134 KH 13.
FT DOMAIN 1135 1209 KH 14.
FT MOD_RES 295 295 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 296 296 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 1268 AA; 141439 MW; 374ECB83D13A7431 CRC64;

Query Match 7.0%; Score 207.5; DB 1; Length 1268;
Best Local Similarity 23.9%; Pred. No. 3.6e-05;
Matches 95; Conservative 69; Mismatches 153; Indels 81; Gaps 18;

QY 202 VPTQVGAIIKEGATIRNITKQTSKIDVHRKENAGAAEKSTITLSTPEGTSAAACKSIL 261
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
157 IPKEHREIVGKNGEKLQDLKATKTIQIPRPD---PSNQIKITGKEGIEKARHEVL 213
QY 262 EIMHKEADIKFTEIPLKILAHNNFVG---RLIKEGRNKKIEQDQDTTKITISPLQEL 318
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 LI--SAEQDKRAVERLEVEKAFHPFIAGPNRLVG-----EIMQETGTRINIPP---- 260

QY 319 TLYNPERT-ITVKGWVEYCAKAEIEMKKIRESYENDIASMNL-----QAHLIPGLNLNA 372
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
261 --PSVNRTEIVFTGKEQLAQAVARI-KKIYEKKKKTTTIAVEKKSHQKVIQPKGNS 317

QY 373 LGLFPPTSG---MPPTS-----GPP-----SAMTPPYPPQFQSETVY-----HQ 410
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
318 LOEILERTGVSEIIPPSDISSETVILRGEPEKLGQALTEVYAKANSFTVSSVAAPSWLHR 377

QY 411 FIPALSVGAIIKQKQHIKQLSRFAGASIKTAPAEAPDAKVMVLIITGPPEAQFKAQGR 470
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
378 F-----ITGKGNLAKITQ-QMPKVHIEFTGED-----KITLEGPTEDVNVAAEQEI 424

QY 471 YGKIKE-----ENFVSPKEVKLEAHIRVPSFAAGRVICKGKGTVNELONLSSAEVVVPRD 526
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 EGWVADLINRMDYVEINIDHKFHRHL-----IGKSGANINRIKQYKVSVRIP--- 472

QY 527 QTPDENDQVVKITGHFYACQVAQRKIQEILTVQVKOHQ 564
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
473 --PDSEKSNLIRIEGDPQGVQQAQKRELLELASRMENER 508

RESULT 10
ID PBP2_YEAST STANDARD; PRT; 413 AA.
AC P38151.
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PAB1-BINDING PROTEIN 2.
GN PBP2 OR YBR233W OR YBR1531.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C.
RA Dubois E., el Bakkoury M., Glansdorff N., Messenguy F., Pierard A.,
RA Scherens B., Vierdeels F.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP PARTIAL CHARACTERIZATION.
RX MEDLINE=99038243; PubMed=9819425;
RA Mangus D.A., Amrani N., Jacobson A.;
RT "Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-
binding protein, regulates polyadenylation.";
RL MOL. CELL. BIOL. 18:7383-7396(1998).
CC -! SUBUNIT: INTERACTS WITH PAB1.
CC -! SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -! SIMILARITY: CONTAINS 3 KH DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z36101; CAA85196.1; -.
DR PIR; S46109; S46109.
DR SGD; S0000437; PBP2.
DR InterPro; IPR000958; KH.
DR Pfam; PF00013; KH-domain; 3.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS00084; KH_TYPE_1; 3.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
FT DOMAIN 66 130 KH 1.
FT DOMAIN 148 213 KH 2.
FT DOMAIN 330 394 KH 3.
SQ SEQUENCE 413 AA; 45782 MW; 638846509BCE1840 CRC64;

Query Match 6.9%; Score 204; DB 1; Length 413;
Best Local Similarity 21.3%; Pred. No. 1.2e-05;
Matches 93; Conservative 78; Mismatches 164; Indels 102; Gaps 17;

QY 155 PDMAAQQNPLOQPRGR-----RGLQGRGSSRQSGVSKOKPCDLPL 198
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 PTTVLVSPNTLRKGGEDTSEQLEAEIKRVALKDADSHDNDHSDPNV---PSDVHL 69

QY 199 RLLVPTQFVGAIIKEGATIRNITKQTSKIDVHRKENAGAAEKSTITLSTPEGTSAAACK 258
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
70 RMLCLVKHASLIVGHKGATISRIKSETSARINISNNIR-GVPERIVYVVRGTCDDVAKAYG 128

QY 259 SILETMHKE-----AODIKFTEEIPLKLAHNNFVRLIGKEGRNLKKIEQDQDTTKITI 312
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 MIVRALLEEHEGHNEDGEDI-----ETISNLIIPHLMGLCIIGRGRSLREIEDLSAAKLFA 184

QY 313 SPLQELTYNPERTITVKGWVEYCAKAEIEMKKIRESYENDIASMNLQAHLIPGLNLNA 372
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 SPNQ--LLLSNDRILTINGVDPDAIHATFYISQTL-----LNFQWE-SPQKNVNR 231

QY 373 LGLFPPTSGMPPTSGPPSAMTPPYQF-----EQSETETV-----HQFIPALSVGAII 422
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
232 SIYYQPT-----QFNSVLIDHSQPNITIFHQRNHQVHP--SDRLLSY 270

QY 423 KQGHKQLS-----RFAGASIKIAPAEAPDAKV-RMVLITGPPEAQFKAQGRYIK 473
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
271 KPNKNLPISSTLLSMATPOYTASVANATAFQFNPVNPVLDGVPVISPAGNHLIMNF 330

QY 474 IKEENFVSPKEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVVPRDQTPDE-N 532
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
331 VQGEIFIDEK-----FVGNVIGDKGHINSVKESTGCSIII---QDPVEGS 373

QY 533 DQVVKITGHFYACQVA 549
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 SERRLTIRGTETMASQAA 390

RESULT 11
PCB2_HUMAN
ID PCB2_HUMAN PRT; 365 AA.
AC Q15366;
```


Query Match 6.8%; Score 202; DB 1; Length 362;

Best Local Similarity 24.2%; Pred. No. 1.3e-05;

Matches 87; Conservative 60; Mismatches 117; Indels 96; Gaps 14;

QY 196 LPLRLVPTQFVGAIGEGATIRNITQTQSKIDVHRKENAGAAEKSTITLSTPEGTS 255

14 LTRILLMHGKVGSIIGKSGESVKMRRESGARINI-----SEGNCPERITLAGP--TNA 67

QY 256 ACKSILETMHKEAQDKFT-----EEIPLKLIAHNNFVGRLLIGKEGRNLKIEQDT 306

68 IFKAFAMIIDKLEEDISSMTWNTASTRPPVTLRLVPASQCSGLIGGCKIKIREST 127

QY 307 DTKITISPLQELTYNPTITVKG---NVETCAKAAEEIMKKI----- 347

128 GAQVQVA--GDMPLPNSTERAITIAGIPQSIIECVKQICVMLESPPKGVTIYRKPSS 185

QY 348 -----RESYENDIASMNLQAHILPGLNLNALGLFPPTSGMP----- 383

186 PVIFAGGQDRYSTGSDASF---PHTTSMCLN-----PDLEGPPLAEAVTIQGYAIPQ 237

QY 384 -----PPTSGPP--SAMTPPYQ-----PEQSETEVYHQF--IPALSVG 418

238 DITKLHLQAMQOSHFPMTHTGTFSGIESSPEVGYNAGLDASQOTTSHELTIIPNDLIG 297

QY 419 AIIQKQGHKIKOLSIFAGASIKIA--PAEAPDAKVRVITGPPEAQFKAQGRIVYKIEE 477

298 CIIHQGAKINEIROMSGAQIKIANPVEG--STDROVITGSAASISLAQYILNVLRSSE 355

RESULT 13

VLGN_CHICK STANDARD; PRT; 1270 AA.

AC P81021;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE VIGILIN.

GN HDLBP OR VGL.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryonic sternal cartilage;

RX MEDLINE=92298988; PubMed=1606952;

RA Schmidt C., Henkel B., Poeschl E., Zorbas H., Purschke W.G.,

RA Gloc T.R., Mueller P.K.;

RT "Complete cDNA sequence of chicken vigilin, a novel protein with

amplified and evolutionary conserved domains."

RL Eur. J. Biochem. 206:625-634(1992).

CC -!- SUBCELLULAR LOCATION: INTRACELLULAR.

CC -!- SIMILARITY: CONTAINS 14 KH DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X65292; CAA46387.1; .

DR InterPro; IPR000958; KH.

DR Pfam; PF00013; KH-domain; 14.

DR SMART; SM00322; KH; 14.

DR PROSITE; PS50084; KH_TYPE_1; 14.

KW RNA-binding; Repeat.

FT DOMAIN 150 188 KH 1.

FT DOMAIN 219 260 KH 2.

FT DOMAIN 291 333 KH 3.
FT DOMAIN 360 402 KH 4.
FT DOMAIN 431 473 KH 5.
FT DOMAIN 504 545 KH 6.
FT DOMAIN 577 619 KH 7.
FT DOMAIN 651 693 KH 8.
FT DOMAIN 724 766 KH 9.
FT DOMAIN 798 840 KH 10.
FT DOMAIN 872 913 KH 11.
FT DOMAIN 970 1012 KH 12.
FT DOMAIN 1051 1093 KH 13.
FT DOMAIN 1126 1168 KH 14.
SQ SEQUENCE 1270 AA; 142220 MW; 0A8844F91F9B8619 CRC64;

Query Match 6.6%; Score 194.5; DB 1; Length 1270;

Best Local Similarity 21.8%; Pred. No. 0.00023;

Matches 97; Conservative 85; Mismatches 168; Indels 95; Gaps 17;

QY 182 QGSGSV--SKQPCDPLR-----LLVPTQFVGAIGEGATIRNITKQTSKI-- 229

483 EGDPOGVQQAQKELLELSRMENERTKOLIEQKPHRTIIGKGERIREIREKFPVEIIN 542

QY 230 ---DVHRKENAGAAEKSTITLSTPEGTSACKSILIMHKEAQDKFTTEIPLKILAHNN 286

543 FPDPAHKSD-----IVQLRGPKNVEKCTKYMOKMVADLVENSFSISVPFIKQFHN 594

QY 287 FVGRLLIGKEGRNLKIEQDTDKITISPLQELTYNPTITVKGNVETCAKAAEEIMKK 346

595 ----IIGKGANIKKIREESNTKIDLPGREQAT----QOLLSQGREQTVKLLRHRL-- 644

QY 347 IRESYENDIASMNLQAHILPGLNLNAL---GLFP-----PTSG----- 381

645 ---AQKELANITEVEVSIPLSLHNSLIGAKGRFIRSIIMECGGVHIFTEGSSATVT 701

QY 382 ---MPPPTSGPPSAMTPPYQFQSOSETV-----HOFIPALSVAIGIGKQGHKQL 431

702 IRAQPTWRKPRSSCCTWAEKQTKSYTVDLRAKPEYHKF-----LIGKGGNIRKV 753

QY 432 SRFAGASIKIAPAEAPDAKVRVITGPPEAQFKAQGRIVYKIE--ENFVSPKEEVKLEA 490

754 RNTGARIIFPTSEDKQDE--LITIMGTEAVKEAKEALEALIKNLNDNVVEDSMVDPKH 811

QY 491 HIRVPSFAAGRVIGKGGKTVNELQNLSAEVYVPRDQTPDENQVYVVKITGHFACQVAQ 550

812 H-----RHEVIRGQVLRLEIADEYGVVMVRLPTVSGTQSDKVTLK--GAKDCVEAAK 861

QY 551 RKIQEILTQVKHQHQQKALQSGPPQ 575

862 KRQIEIIDL---EAQVTIETIPQ 883

RESULT 14

ROK_HUMAN

ID ROK_HUMAN STANDARD; PRT; 463 AA.

AC Q07244; Q15671;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K (HNRNP K) (DC-STRETCH

BINDING PROTEIN) (CSBP) (TRANSFORMATION UPREGULATED NUCLEAR PROTEIN)

DE (TUNP).

GN HNRPK OR HNRNPK.

OS Homo sapiens (Human), and Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606, 10116;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=Human.

RX MEDLINE=92107165; PubMed=1729596;

RA Matunis M.J., Michael W.M., Dreyfuss G.;

RT "Characterization and primary structure of the poly(C)-binding

Query Match 6.4%; Score 190; DB 1; Length 463;
Best Local Similarity 20.7%; Pred. NO. 0.0001;
Matches 87; Conservative 54; Mismatches 131; Indels 148; Gaps 13;

Search completed: January 25, 2002, 22:20:22
Job time: 454 sec

